A Video Processing Based System for Counting Vehicles

Mr. Mahalingesh Naduvinamani[1], Prof. S A Angadi[2]

[1]Student Department of Studies in Computer Science & Engineering, VTU Belagavi, India 590018
[2]Head of Department of Computer Science & Engineering, VTU Belagavi, India 590018

Abstract - *Counting of vehicles passing a particular point is of interest to town/traffic planners. In this paper a new technique for counting of vehicles passing a particular location is described. The proposed methodology processes a video, that captures all the vehicles passing through a point, breaks it down to individual frames and it counts the moving vehicles passing through the point by recognizing a vehicle in previously marked position. The Discrete Wavelet Transform (DWT) features and neural networks are employed for the task. The results obtained on the videos tested are satisfactory.*

***Key Words*: DWT features, vehicle counting, neural networks.**

1. INTRODUCTION

Video processing systems require a stream processing architecture in which video frames from a continuous stream are processed one (or more) at a time. This type of processing is critical in systems that have live video or where the video data is so large that loading the entire set into the workspace is inefficient. Computer Vision System Toolbox supports a stream processing architecture through System objects and blocks.

In recent years, as a result of the increase in vehicle traffic, many problems have appeared. For example, traffic accidents, traffic congestion, traffic induced air pollution and so on. Traffic congestion has been a significantly challenging problem. It has widely been realized that increase of preliminary transportation infrastructure e.g., more pavements, and widened road, have not been able to relieve city congestion. As a result, many investigators have paid their attentions on intelligent transportation system (ITS), such as predict the traffic flow on the basis of monitoring the activities at traffic intersections for detecting congestions. To better understand traffic flow, an increasing reliance on traffic surveillance is necessary. Automatic detection of vehicles in video surveillance data is a very challenging problem in computer vision with important practical applications, such as traffic analysis and security. Vehicle detection and counting is important in computing traffic congestion on highways. A system like the one proposed here can provide important data for designing roads and utilities.

1.1 Brief Survey

A brief survey of the related work in the area of video segmentation and traffic surveillance is presented in the following. Chen et al., [1], [2] have addressed the issues regarding unsupervised image segmentation and object modeling with multimedia inputs to capture the spatial and temporal behavior of the object for traffic monitoring. In [3] algorithms for vision-based detection and classification of vehicles in monocular image sequences of traffic scenes are recorded by a stationary camera.

Fig1. Detected vehicles

Processing is done at three levels: raw images, region level, and vehicle level. Vehicles are modeled as rectangular patterns with certain dynamic behavior. Daniel et al., [4] presents the background subtraction and modeling technique that estimates the traffic speed using a sequence of images from an un calibrated camera. The combination of moving cameras and lack of calibration makes the concept of speed estimation a challenging job. Cheng and Kamath [5] compare the performance of a large set of different background models on urban traffic video. They experimented with sequences filmed in weather conditions such as snow and fog, for which a robust background model is required. Kanhere et al., [6] applies a feature tracking approach to traffic viewed from a low-angle off axis camera. Vehicle occlusions and perspective effects pose a more significant challenge for a camera placed low to the ground. Deva et al., [7] propose a concept to automatically track the articulations of people from video sequences. This is a challenging task but contains a rich body of relevant literature. It can identify and track individuals and count distinct people. Toufiqs P. et al., in [8] describes background subtraction as the widely used paradigm for detection of moving objects in videos taken from static camera which has a very wide range of applications. The main idea behind this concept is to automatically generate and maintain a representation of the background, which can be later used to classify any new observation as background or foreground. In [9] background subtraction also involves computing a reference image and subtracting each new frame from this image and thresholding the result. This method is an improved version of adaptive background mixture model, it is faster and adapts effectively to changing environments.

The proposed methodology used the DWT features of the segmented image portion containing vehicles to train a back propagation Neural Network to detect vehicles. Further during testing the video is segmented into frames and the location of frame, which depicts the fixed point is used to extract the DWT feature which are input to the back propagation Neural Network. The back propagation Neural Network detects a vehicle when one such vehicle is present in the location of the image. The proposed methodology has been tested on a few videos and the results are satisfactory.

The rest of the paper is organized into three sections. Section II describes the proposed methodology. Section III presents the experimentation and results. Finally section IV gives the conclusion.

2. PROPOSED METHODOLOGY

The block diagram of the proposed methodology is shown in figure 2. The methodology is made up of two phases, namely training and testing phase. The vehicle images in the videos are initially manually extracted and used to train a back propagation neural networks (BPNN) for vehicle detection DWT features. In the testing phase the video containing vehicle moving across a given location the pre-process to extract frames and each of the frame is process to extract the DWT features, which are fed to the neural networks (BPNN). The neural network then detects the existence the vehicles or otherwise which is then counted and the results are displayed.

The Video Processing Based System for Counting Vehicles consists of three modules:
Module 1: Pre-processing the images (video to frames conversion)
Module 2: DWT object (feature) extraction
Module 3: Neural Networks training method.

2.1 VIDEO TO FRAMES CONVERSION

A Video Processing Based System for Counting Vehicles is to collect a database of road videos. First step is video is converted into number of frames. The video must be in .mp4 format. Video to frame conversion using the function mmreader, those converted frames are in .jpg format and are stored in database.

Those images are sorted or filtered and further given to the feature extraction module. Those images are then trained using the extraction features. Further the same steps are also used in testing.

2.2 DWT FEATURE EXTRACTION METHOD

In Video Processing Based System for Counting Vehicles the second step is extraction using 2D DWT feature extraction method.

The two dimensional DWT (2D-DWT) is nowadays established as a key operation in image processing. It is multi-resolution analysis and it decomposes images into wavelet coefficients and scaling function. In Discrete Wavelet Transform, signal energy concentrates to specific wavelet coefficients. This characteristic is useful for compressing images. Apply first level decomposition and decompose image in 4 subparts i.e. HH, HL, LL and LH.

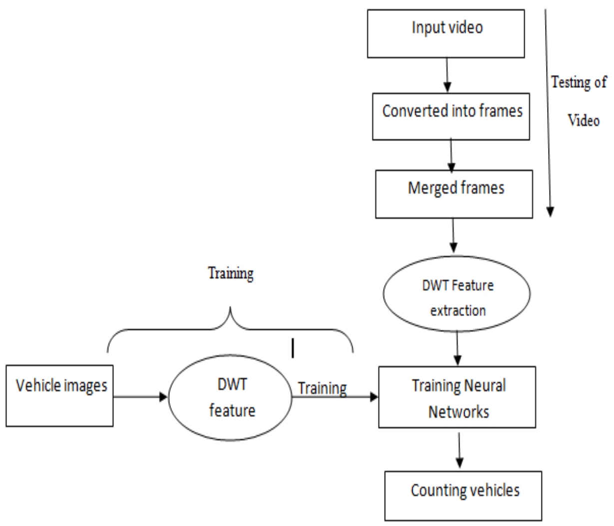

Fig2. Architecture of a Video Processing Based System for Counting Vehicles

Discrete wavelet transform (DWT) is a well-known signal processing field tool; it is widely used in feature extraction and compression and de-noising applications. The discrete wavelet transform has been used in various feature extraction studies.

All frames in the video are transformed to the wavelet domain. The frames are decomposed in 1st level sub band frames by separable 2- D wavelet transform. The Discrete Wavelet Transform, decompose a signal into sub bands with smaller bandwidths and slower sample rates. For this purpose it uses a filter bank with low pass and high pass FIR filters, that is Dyadic Filter, that we specify either directly on the block mask, or using wavelets from the Wavelet Toolbox. The low pass and high pass filters are usually half-band filters designed to complement each other. The representation of an image I after 1-level DWT with its sub-bands is given by the Eq. 1

$$I = I_a^1 + \{I_h^1 + I_v^1 + I_d^1\} \quad (1)$$

where I_a^1 represents the approximation of input image (smaller scaled form) and h , I_h^1 , I_v^1 , I_d^1 represent horizontal, vertical and diagonal details respectively, where the powers of the terms represent the level of decomposition. Further decompositions can be achieved by decomposing the LL sub band successively and the resultant image is split into multiple bands.

2- Level DWT decomposition. Highest level of decomposition depends upon the wavelet filter used, need of the application and features required for the classification. Using 4th level of DWT decomposition, coefficients of approximate & detail Sub-bands are extracted. Based on the available wavelet coefficients, Eq. 2 Energy and standard deviation is capable. In DWT feature extraction HAAR wavelet is used.

$$E_k = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|x_k(i,j)| \quad (2)$$

2.3 USING NEURAL NETWORKS TRAINING METHOD

An artificial neural network (ANN) [4], often called as a "neural network" (NN), is a computational model based on the biological neural networks, in other words, it is a representation and emulation of human neural system. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In practical terms neural networks are non-linear statistical data modelling tools.

Back Propagation network is considered to be quintessential Neural Network. Back Propagation is the training or learning algorithm rather than the network itself. To train the network we need to give the output called the Target for a particular input. The input and its corresponding target are called a Training Pair. Once the network is trained, it will provide the desired output for any of the input patterns. The network is first initialized by setting up all its weights to be small random numbers – say between –1 and +1. Next, the input pattern is applied and the output is calculated this is called the forward pass. The calculation gives an output which is completely different to what is expected (the Target), since all the weights are random. We then calculate the Error of each neuron, which is essentially: Target - Actual Output. This error is then used mathematically to change the weights in such a way that the error will get smaller. In other words, the Output of each neuron will get closer to its Target (this part is called the reverse pass). The process is repeated again and again until the error is minimal. In the proposed methodology the number of input (feature) is input-4output-1. The middle level neurons are 3. The testing and experimentation of the methodology is given in section III.

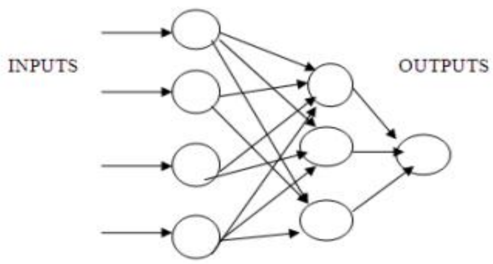

Fig3. Inputs and outputs of the neural network

3. EXPERIMENTATION

In the proposed methodology the first step is to input the video into the mat lab tool. The images of the vehicles are used for training the neural network (back propagation methodology) tool. The GUI of the system is developed which is shown in fig 4. Figure5 shows the feature extraction and frame caption.

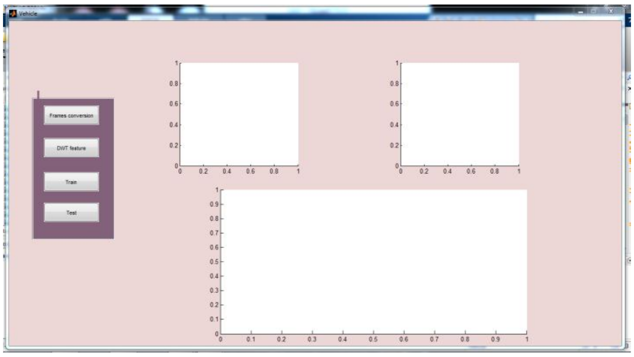

Fig4. Result of GUI to the user

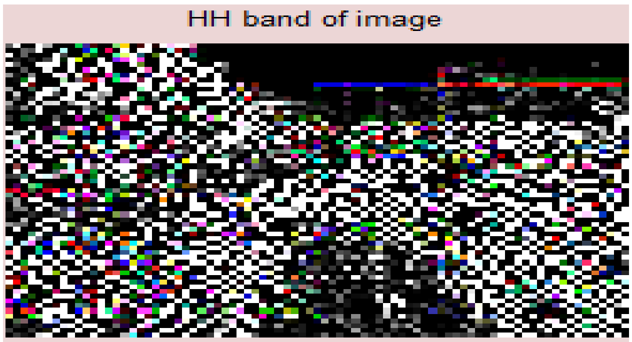

Fig5. Result of DWT feature extraction frame

Fig6. Before detected vehicles

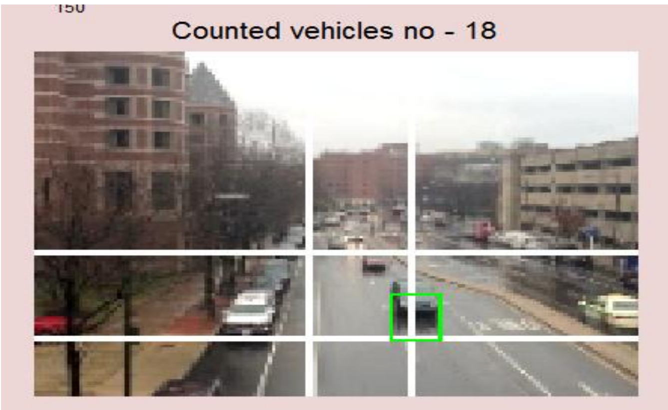

Fig7. Result of finally detected and counted vehicles in video

.Further the video before and detection and counting through Neural Network training is shown in figure 6 and 7. The methodology is tested thoroughly and the results are satisfying.

Experimental results are shown in Table I. The accuracy of the proposed vehicle counting method varied from 97.9 to 100%, depending on the input videos. It suggests that the proposed method could perform quite well on every tested video.

TABLE 1. EXPERIMENT RESULTS

Video number	Total number of vehicles	Number of counted vehicles	Accuracy (%)
Vid 1	**48**	**47**	**97.91**
Vid 2	**18**	**18**	**100**

4. CONCLUSION

In this paper, we present a system that has been developed to detect and count the objects in a video efficiently. The system effectively combines simple domain knowledge about object classes to identify target objects in the presence of partial occlusions and ambiguous poses, and the background clutter is effectively rejected. The experimental results show the accuracy of counting vehicles was full, although the vehicle detection was 100% which is attributed towards partial occlusions. The computational complexity of our algorithm is linear in the size of a video frame and the number of vehicles detected. As we have considered traffic on highways there is no question of shadow of objects such as trees, a but sometimes due to occlusions two objects are merged together and treated as a single entity.

In this methodology accurate results are shown. We mainly use DWT feature for training the objects and Neural Networks for matching the trained objects into input video.

REFERENCES

[1] Chen S. C., Shyu M. L. and Zhang C., "An Unsupervised Segmentation Framework for Texture Image Queries", In the 25th IEEE Computer Society International Computer Software and Applications Conference (COMPSAC), Chicago, Illinois, USA, Oct.2000.

[2] Chen S. C., Shyu M. L. and Zhang C., "An Intelligent Framework for Spatio-Temporal Vehicle Tracking", 4th International IEEE Conference on Intelligent Transportation Systems, Oakland,California, USA, Aug. 2001

[3] Gupte S., Masoud O., Martin R. F. K. and Papanikolopoulos N. P.,"Detection and Classification of Vehicles", In IEEE Transactions on Intelligent Transportation Systems, vol. 3, no. 1, March, 2002, pp. 37– 47.

[4] Dailey D. J., Cathey F. and Pumrin S., "An Algorithm to Estimate Mean Traffic Speed Using Uncalibrated Cameras", In IEEE Transactions on Intelligent Ttransportations Systems, vol. 1, no. 2,pp. 98-107, June, 2000.

[5] S. Cheung and C. Kamath, "Robust Techniques for BackgroundSubtraction in Urban Traffic Video", In Video Communications and Image Processing, SPIE Electronic Imaging, San Jose, January, 2004.

[6] N. Kanhere, S. Pundlik and S. Birchfield, "Vehicle Segmentation and Tracking from a Low-Angle Off-Axis Camera", In IEEE Conference on Computer Vision and Pattern Recognition", San Diego, June,2005.

[7] Deva R., David A., Forsyth and Andrew Z., " Tracking People by Learning their Appearance", In IEEE

Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no1, Jan. 2007.

[8] Toufiq P., Ahmed Egammal and Anurag Mittal, "A Framework for Feature Selection for Background Subtraction", In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), 2006.

[9] P. Kaewtra Kulpong and R. Bowden, "An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection", In Proceedings of the 2nd European Workshop on Advanced Video-Based Surveillance Systems, Sept. 2001.

BIOGRAPHIES

Prof. Shanmukappa. A. Angadi
Is a professor and Head of the Department of computer science and Engineering, center for PG Studies, Visvesvarayya Technological University, Belagavi Karnataka, India.
His research areas are Image processing and pattern recognition, Soft computing, Internet of things (IOT) and Graph Theoretic Technologies.

Mr.Mahalingesh Naduvinamani
Student
Department of Studies in Computer Science & Engineering, VTU Belagavi, India 590018