Traffic flow characterization is critical for transportation systems and urban infrastructure planning. In most major cities in Latin America, this characterization is carried out visually by a group of observers during one-hour sampling periods. Automating the traffic flow characterization should result in a number of benefits including higher consistency and reliability in the output statistics, as well as keeping the people in charge of this process safe. In this project, an automatic vehicle counting and classification system prototype is presented. The system uses video images previously recorded thru a camcorder and, upon an initial setup session, proves capable of detecting, counting and classifying the passing vehicles to an acceptable error rate. The results of testing the prototype in a video sequence of 90 minutes respectively are presented. The preliminary evaluation is promising and provides a solid base for the development of the complete system.

Introduction

Several key decisions affecting areas from traffic control to urban infrastructure have to do with an accurate account of vehicles in particular points in a city. The required traffic flow characterization process can mainly be broken down into three components: counting the number of vehicles, classifying these vehicles, and determining which direction these vehicles followed. Currently, these three activities are carried out mostly through visual means by groups of observers positioned around the areas of interest in all major cities in Latin America. The characterization process is usually carried out in 1-hour periods during characteristic times during the day. Needless to say, the
characterization through this means is hardly consistent due to its inherent human limitations like attention span or visual obstruction by large vehicles.

An automated system for traffic flow characterization was set to be developed with the idea that it utilized methods nonintrusive to the pavement and that it could be easy to install. The existence of video cameras already installed around the metropolitan area where this development took place -Monterrey, México- made such technology the ideal candidate for its adoption. Following this decision, it became necessary to separate the development of the system in two phases: the first phase included counting and classification and the other included the determination of vehicle directions. Counting could then be approached through video detection and classifying through artificial neural networks. The second phase is currently on the works by our research group.

The construction of the system prototype is described in this paper, making emphasis on the evaluation results which are believed to provide a solid base for developing the complete system.

Background

Traffic flow characterization has been discussed in the literature since the 70’s (Sun and Ritchie, 2003). For vehicle detection and tracking, several methods have been utilized ranging from inductive loops (Pursula and Pikkarainen, 1994; Sun and Ritchie, 2000; Oh et al., 2002), piezoelectric sensors (Li et al., 2006), magnetometers (Cheung et al. 2004), infrared and laser sensors (Harlow and Peng, 2001; Hussain and Moussa, 2005), acoustic sensors (Nooralahiyan and Kirby, 1998; Duarte and Hu, 2004), microwaves (Urazghildiiev et al., 2002), all the way to computer vision (Wei et al., 1996; Ma and Grimson, 2005; Mei et al., 2006). According to Gupte, et al. (2002), in spite of the existence of several works regarding vehicle detection and tracking, there has been little attention paid to the issue of vehicle classification.

The use of computer vision for vehicle detection and count seemed to match the desired quality of not being an intrusive method. Furthermore, this technology had been proved able to provide information with regards to the length of a vehicle queue, traffic speed, as well as directional distribution (Ha et al., 2004).
Artificial Neural Networks have been known for showing a competitive performance in classification problems (Hagan et al., 1995). This has also been verified for traffic flow (Wu et al., 2001; Ha et al., 2004).

Based on these pieces of information, it was decided to couple the use of computer vision and artificial neural networks for building the intended system.

**Proposed Integration of Techniques**

Two phases integrate the task to characterize the traffic flow in this work: vehicle detection/counting and vehicle classification. The former is used to estimate the features of particular vehicles and count vehicle occurrences, and the latter provides the adequate processing to these features to determine how a particular vehicle should be classified.

Vehicle detection and counting is carried out by means of computer vision. The electronic images were taken using a camcorder mounted on a previously selected intersection of the metropolitan area of Monterrey. In a normal intended operation of the system the information flow would entail detecting the vehicles and extracting their features, to then feed these features into an artificial neural network for classification. This flow is schematically shown in Figure 1. Although it is intended for the system to work online in the future, the work reported up to this point considers preredcorded video sequences. In addition, a training period for the artificial neural network is necessary as it will be explained later.
Three different classes were defined for this project: (1) small vehicles, (2) medium-sized vehicles, and (3) large vehicles. The first class is conformed by all vehicles identified by their manufacturers as light vehicles. The second class includes vans, SUVs, pick-up trucks and the like. The third class includes all transportation buses, trailers and semi-trailers.

The classification scheme obeys to the relationship between the dimensions of a vehicle, its mass, and its impact on pavement wear. It is a well known fact that most of the methods to determine pavement requirements make use of estimates of the number of vehicles and their types that will use a particular road (Gardner, 2000) to then estimate the expected mass and the expected pavement wear. The Mexican Institute of Transportation reports a direct relationship between the decay of pavement and the physical features of the vehicles moving on top of it, the speed at which these move, their load sizes, as well as the characterisitcs of their suspensions’ rigidity and damping properties (Lozano Guzmán et al., 1999).

**Image Processing**

1) Record Video Sequence

The first step includes recording the video sequences of interest. Cameras already installed in different points of the metropolitan area were used for this purpose.
2) Digitalize the Video Sequence

Digitalizing and compressing the video sequences help the purpose of creating a computer file for its analysis. In a future application these processes could be done online, as opposed to the offline scheme adopted up to this point.

3) Process Image

Since a video sequence is made of several static frames that are projected at small time intervals (1/24, 1/35 or 1/30 of a second), each of these images can be analyzed separately. Static image processing makes use of segmentation techniques such as edge detection and background estimation to identify the vehicles present at the image of interest [REFERENCE]. Usually segmentation is not applied to all images in a video sequence, but to those images at every certain number of positions previously determined by the analyst.

4) Feature extraction

The idea behind this step is characterizing a vehicle to be correctly categorized based on the known features of other similar vehicles (Duda et al., 2001). This idea leads to trying to find features that are easy to obtain and that are invariant under several types of noise.

Bishop (1995) states that geometric parameters such as the edge length or the area can be the most appropriate for objects in an image. With this in mind, a series of features for the detected objects (vehicles) in binary format were calculated. These properties of the identified object, all of them measured in pixels, were (1) area, (2) length of the major axis of the circumscribing ellipse, (3) length of the minor axis of the circumscribing ellipse, and (4) perimeter.

The length and the width of a particular vehicle were estimated using (2) and (3) as surrogate measures. All four features were used to describe a given vehicle. These features could then be used as inputs to a classification neural network.
Figure 2 shows an example of vehicle detection for each of the vehicle classes used in this study.

5) Vehicle Counting

Once all the vehicles have been detected in a video sequence, vehicle counting comes out simply as adding all the identification occurrences. Vehicle detection and counting are tightly coupled this way.
**Vehicle classification through Artificial Neural Networks**

1) Define RNA architecture

An Artificial Neural Network similar to the one shown in Figure 3 was used as the classifier in this work. The ANN has three neuron layers: (i) an input layer, (ii) a hidden layer, and (iii) an output layer. Layer (i) receives the vehicle features previously extracted; layer (ii) establishes the classification frontiers and layer (iii) will output the class to which the identified vehicle belongs (small, medium-sized, or large).

![Artificial Neural Network](image)

**Figure 3. Artificial Neural Network used in this work.**

2) Define training patterns

In order for the ANN to be usable, it must have been trained with known patterns from each intended class. The user must define these patterns by visually identifying the vehicles in a segment of the video and indicating to which class
each vehicle belongs. The more patterns available the better classification rate that can be expected from the ANN. All the available patterns must be separated in three disjoint sets: a training set, a validation set, and a test set.

During training, the patterns in the associated set are presented to the ANN for it to adjust its fitting parameters through a variation of the backpropagation algorithm. The aim is to classify all training patterns as accurately as possible by minimizing the classification error throughout the entire set. A thorough treatment on the backpropagation algorithm and its variations can be found in Hagan et al., 1995. The validation set is used for keeping the ANN from overtraining i.e. loosing its generalization capability.

3) Preprocessing information

It is a common practice to preprocess information before using it in an ANN due to, among other things, its effect on the generalization capability (Bishop, 1995). Other advantages of this practice have to do with reducing the effect of having different scales in the inputs (Hansen & Nelson, 2003).

4) Testing the classifier

Because both the training and the validation sets are used to create the ANN, it would not be objective to test the ANN’s classification performance using either of these. Instead the test set is used for assessment purposes.

5) Classification

Given a calibration for the camera as well as an intersection or street to be analyzed, the user must train the ANN once under these conditions for a given period of time. From then on, using exactly the same conditions, the system can use the ANN to classify vehicles for longer periods of time.
Results

The results of using the prototype to analyze a video sequence are presented in this section. The video sequence is ninety minutes long and corresponds to a realistic scenario in which the system will be used. The video was digitalized without sound to a resolution of 352 x 240 pixels in AVI format at a rate of 30 frames/second. The camera used was an Elbex EX/C100/6 installed in a known intersection in the metropolitan area. The analysis was run for every 15th frame A visual vehicle count and classification were performed for comparison purposes.

Because the classification schemes used by other authors are different from that presented in this work, a direct comparison was not possible.

Vehicle Detection and Counting

All vehicles were counted and classified visually for comparison purposes. The video had to be broken down into three segments of half an hour each for convenience of computer processing. In the first, second and third half hours, the vehicle counts were 1625, 1572 and 1671 respectively. Table 1 shows the results of the automatic system vs. the visual count.

Table 1. Comparison of the visual count vs. the automatic count in the 90-minute video.

<table>
<thead>
<tr>
<th>Class</th>
<th>Visual Count</th>
<th>Automatic Count</th>
<th>Automatic Count/ Visual Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>3,076</td>
<td>3,146</td>
<td>102.27%</td>
</tr>
<tr>
<td>Medium-Sized</td>
<td>1,811</td>
<td>1,700</td>
<td>93.87%</td>
</tr>
<tr>
<td>Large</td>
<td>62</td>
<td>22</td>
<td>35.48%</td>
</tr>
<tr>
<td>Total</td>
<td>4,949</td>
<td>4,868</td>
<td></td>
</tr>
</tbody>
</table>

According to Table 1, more competitive detection/counting rates were obtained in this case for the medium-sized, being the small vehicles still the ones with the best rate. The lowest rate was for the large vehicles due to the difficulty of not being able to keep them contained within the detection area. Although it might sound tempting to adjust the detection area, this would conflict with the detection rates of the other classes, making them more prone to error when tracking them within larger areas.
In spite of the classification rate being competitive, there were some troubles with the vehicle detection and counting tasks. Two main problems were identified: (1) double detection/counting, and (2) no vehicle detection/counting. Additionally, some smaller vehicles that should have been discarded –motorcycles and bicycles- were not. Table 2 shows that, out of the 4868 vehicles detected and counted automatically, 405 were double counted, 35 motorcycles and bicycles were detected and 521 were not detected at all. Subtracting those that were double counted as well as subtracting the motorcycles and bicycles, and adding those that were omitted, the new total is 241, which is equal to that of the visual count.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detected</td>
<td>4868</td>
</tr>
<tr>
<td>Double Counted</td>
<td>405</td>
</tr>
<tr>
<td>Motorcycles and Bicycles</td>
<td>35</td>
</tr>
<tr>
<td>Undetected</td>
<td>521</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4949</strong></td>
</tr>
</tbody>
</table>

The double counting error was due mostly to the frame-to-frame comparison process used for detection purposes. Our observation was that this comparison process is highly influenced by the vehicles’ speed. Vehicles at low speeds have a higher probability of being double detected and therefore, double counted.

The no-detection problem had several causes. These are reported in Table 3: (a) shared pixels, (b) vehicle too close to the street edge, (c) low number of pixels and (d) vehicles too close to each other that were taken as one vehicle.

<table>
<thead>
<tr>
<th>Trouble</th>
<th>Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shared Pixels</td>
<td>325</td>
</tr>
<tr>
<td>Vehicle too close to the street edge</td>
<td>156</td>
</tr>
<tr>
<td>Low number of pixels</td>
<td>17</td>
</tr>
<tr>
<td>Vehicles too close to each other</td>
<td>23</td>
</tr>
</tbody>
</table>

In the 325 cases where no-detection occurred due to shared pixels, the problem was that in a first analyzed frame the vehicle was detected in a particular position and in a second analyzed frame, there was another vehicle occupying most of the same position, and therefore, a large number of pixels. This causes the system to discard the vehicle for considering that it is the same vehicle that had not moved in the frame-to-frame comparison.
Eliminating a vehicle that is too close to the street edge is necessary to avoid partitioning a vehicle that is partially out of the view. However, since the edges of the street are joined to the border of the view finder of the camera, a vehicle can be discriminated if the system considers that it is part of the edge of the street, as it happened in 156 occasions.

The system discriminates all objects with a total area of less than 300 pixels. This value was set as a threshold to avoid counting people, bicycles or other objects that are not vehicles and that could appear in the area of analysis. In the 17 occasions when a vehicle was discriminated, the vehicle was too far and because of the perspective, it occupied less than 300 pixels.

The last detection trouble was the fusion of two vehicles too close to each other into only one vehicle.

One possible solution to the troubles described could be the use of vehicle tracking, which will be attempted in the continuation of this work.

**Vehicle Classification**

In order to determine how to better use the ANN for classification, a design of experiments was set up to determine an adequate size of the training set as well as the number of neurons in the hidden layer. The size of the training set is important, since in order to prepare this set, the user must spend time specifying the patterns. These patterns correspond to the number of vehicles detected during a time interval in the video. Certainly, it is desirable to have enough patterns to guarantee a good performance but not too many as to be impractical for the user. On the other hand, the number of neurons in the hidden layer will determine the complexity of the classification boundaries. A very complex boundary will perform very well with the training patterns, but will very likely perform poorly with other patterns.

Three different training time intervals were defined for the experiment: eight, ten and thirty minutes; and three values for the number of hidden neurons: eight, twelve and sixteen. This way, a full factorial was used i.e. all possible
combinations of both variables at their three levels for a total of $3^2 = 9$ runs. The results are graphically shown in Figure 3.

![Figure 3](image)

**Figure 3.** Response surface of the average percentage of correctly classified vehicles for the 90-minute video

Referring to Figure 3, the average percentage of correctly classified vehicles fell in the range from 75.9% to 76.4%. Additionally, it can be seen that training time’s effect is significant while the number of hidden neurons is not. This is true at a statistical significance level of $\alpha=0.05$. The latter can be explained by the technique used to train the ANN in this case. This technique included an early stopping feature for training that made use of both the training set and the validation set to stop the algorithm when the generalization capability of the ANN started to deteriorate. The behavior of the classification rate as a function of the training time is rather helpful since it indicates a region where it the classification rate becomes insensitive. Such a behavior would allow setting the training time to the lowest value of the stable region. In this particular case, it would have implied using a training time of 20 minutes.
Conclusions

In this work, a system to automatically detect, count and classify vehicles was presented. The system is in its prototypical phase; however, the preliminary results allow evaluating its potential capabilities. In fact, the 90-minute video sequence, is quite a realistic setting. The system is intended to use infrastructure already in place in the metropolitan area of Monterrey and to be the least intrusive possible on the pavement.

The results show that the ANNs are quite competitive as nonlinear classifiers with a comparatively modest number of training patterns, and will therefore be kept as part of the system.

The project was geared, in a first phase of development, to the tasks of vehicle counting and classification. Obviously, the system is still to be improved as well to be extended, in a second phase, to detect the direction of each passing vehicle. Of especial importance will be improving the system to detect, count and correctly classify large vehicles, since these are the ones that cause the most damage to the pavement. Many of the challenges set forth by the preliminary evaluation of the system will be attempted to be solved through the use of vehicle tracking techniques in the future.

References


