Minnesota Department of Transportation

RESEARCH SERVICES

Office of Policy Analysis, Research & Innovation

Practical Methods for Analyzing Pedestrian and Bicycle Use of a Transportation Facility

Nikolaos P. Papanikolopoulos, Principal Investigator
Department of Computer Science and Engineering
University of Minnesota

February 2010

Research Project
Final Report #2010-06

Your Destination...Our Priority

Technical Report Documentation Page

1. Report No. MN/RC 2010-06	2.	3. Recipients Accession No.
4. Title and Subtitle Practical Methods for Analyzing Pedestrian and Bicycle Use of a Transportation Facility		5. Report Date February 2010 6.
7. Author(s) Guruprasad Somasundaram, Vassilios Morellas, Nikolaos P. Papanikolopoulos		8. Performing Organization Report No.
9. Performing Organization Name and Address Department of Computer Science and Engineering University of Minnesota 200 Union St SE Minneapolis, MN 55455		10. Project/Task/Work Unit No. 11. Contract (C) or Grant (G) No. (c) 89261 (wo) 62
12. Sponsoring Organization Name and Address Minnesota Department of Transportation Office of Research Services 395 John Ireland Boulevard, MS 330 St. Paul, MN 55155		13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code
15. Supplementary Notes http://www.lrrb.org/pdf/201006.pdf		

16. Abstract (Limit: 250 words)

The objective of the project is to analyze existing technologies used for the process of generating counts of bicycles and pedestrians in transportation facilities such as walk and bicycle bridges, urban bicycle routes, bicycle trails etc. The advantages and disadvantages of each existing technology which is being applied to counting has been analyzed and some commercially available products were listed. A technical description of different methods that were considered for vision based object recognition is also mentioned along with the reasons as to why such methods were overlooked for our problem. Support Vector Machines were used for classification based on a vocabulary of features built using interest point detectors. After finalizing the software and hardware, five sites were picked for filming and about 10 hours of video was acquired in all. A portion of the video data was used for training and the remainder was used for testing the algorithm's accuracy. Results of counts are provided and an interpretation of these results is provided in this report. Upon detailed analysis the reasons for false counts and undercounting in some cases have been identified and current work concerns dealing with these issues. Changes are being made to the system to improve the accuracy with the current level of training and make the system available for practitioners to perform counting.

17. Document Analysis/Descriptors Bicycles, Pedestrians, Traffic, Traffic counting, Computer vision, Machine learning, Visual model, Traffic models, Validation, Training, Training and cross validation		18. Availability Statement No restrictions. Document available from: National Technical Information Services, Springfield, Virginia 22161	
19. Security Class (this report) Unclassified	20. Security Class (this page) Unclassified	21. No. of Pages 41	22. Price

Practical Methods for Analyzing Pedestrian and Bicycle Use of a Transportation Facility

Final Report

Prepared by:

Guruprasad Somasundaram
Vassilios Morellas
Nikolaos P. Papanikolopoulos

Artificial Intelligence, Robotics and Vision Laboratory
Department of Computer Science and Engineering
University of Minnesota

February 2010

Published by:

Minnesota Department of Transportation
Office of Research Services
395 John Ireland Boulevard, MS 330
St. Paul, Minnesota 55155-1899

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

Pedestrian and bicycle counting has been carried out for a long time. One of the most primitive methods is the turnstile. By forcing pedestrians through turnstiles and counting the number of times the turnstile was used provided a good count of the number of pedestrians. There is also widespread use of manual counts where volunteers sit at a location and make hash-marks on paper. However, with technological advances many different methods of counting pedestrians and bicycles have evolved. Every method has its pros and cons and its effectiveness is good in some situations and worse in others.

This research project looks into vision-based recognition technology and how it may be a viable alternative to many products on the market, including overcoming many of their shortcomings. Also the overall area of object recognition and classification, under which this project can be categorized, is a vastly researched area with a steady growth in terms of findings and accuracy of results. Using such state-of-the-art methods, it is possible to realize a robust vision-based system capable of using a variety of visual cues from traffic videos to detect and successfully count bicycles with an acceptable degree of accuracy.

A novel computer vision based algorithm based on modern machine learning ideas was developed for the recognition of bicycles and pedestrians. This algorithm was tested on standardized datasets and its performance was convincing enough to be tested on practical cases with real life video. After some modifications based on practical observations, the algorithm was implemented and tested on some real life videos. The results were encouraging, and the system proved to be a well-suited prototype for this counting application. Work is in progress to improve this prototype and bring it to the market as a product.

CHAPTER 1

SUMMARY OF EXISTING METHODS FOR BICYCLE AND PEDESTRIAN COUNTING

(Refer Appendix A for tabulation of different methods.)

Buried Pressure Pads

As the name suggests these are pressure pads which are buried underground. Whenever a bicycle passes over the pads the pressure pulses are transmitted as acoustic waves through water (hydro-acoustic) or other material mediums (mechano-acoustic).

Advantages

The manufacturer's claim is that these devices are highly reliable and accurate. These are suitable to be used in trails and can fit any path width. Since they count pulses the devices can count bicycles that pass side by side.

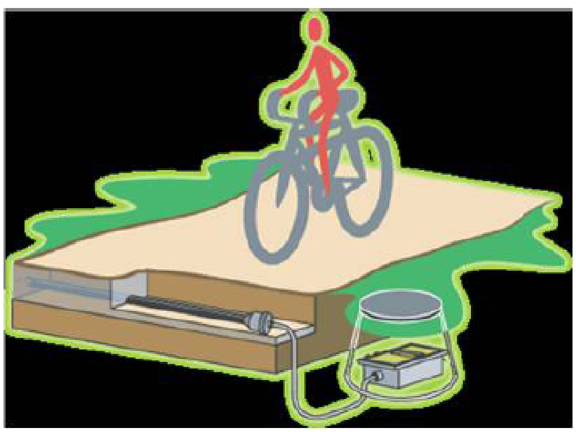

Figure 1: Buried pressure pads.

Disadvantages

However these devices are not completely weather resistant. When the ground is frozen solid the device doesn't function. Also if there are very wide paths it is difficult to count all bicycles. There's also no way of knowing what actually caused the count a bicycle or a pedestrian or something else. Fundamental to the principle is the fact that the bicycle has to directly step over the pressure pads. Further the installation of this device requires expert handling and a lot of time. However the major concern of this device is how fast the device can recover to its initial state after successive stepping over the pads.

Products

Ecocounter and TCS Instruments are some of the manufacturers of buried pressure sensors.

Pyroelectric Sensor

In this method, a lens sensitive to the infrared radiation emitted by the human body detects each time a person passes. The narrowness of the area means that two people following each other closely can be counted.

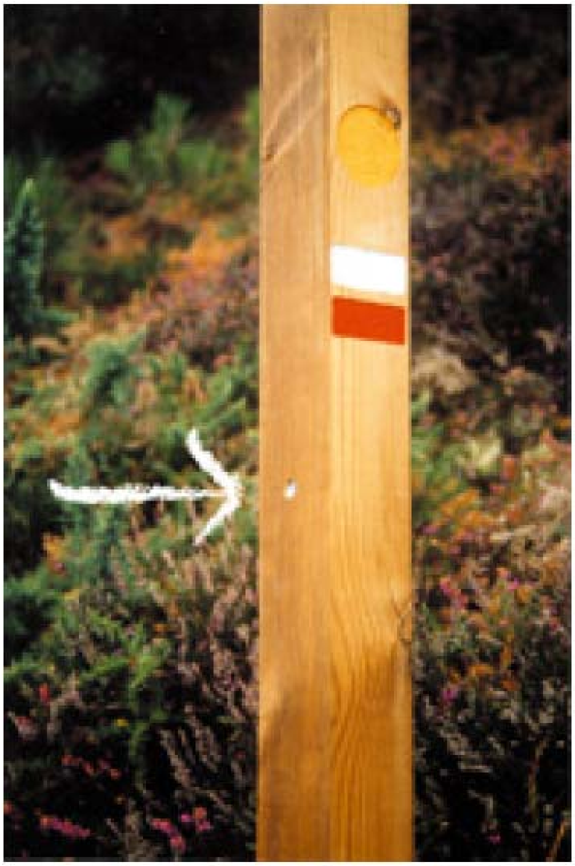

Figure 2: Pyroelectric sensor.

Advantages

These systems are adapted to hard and eroded ground. They can be used for snowy ground too. They are also highly portable as they are compact and light weight.

Disadvantages

Since they use heat to detect bodies of pedestrians, external heat sources present noise. Though they can count pedestrians passing by one after the other, performance when pedestrians passing side-by-side may be hampered.

Products

Ecocounters also manufacture these types of sensors for urban areas.

Figure 3: Inductive loop sensor.

Inductive Loop

Inductive loops are common for vehicle counting. However it can also be used for counting bicycles. The alloys used in bicycles are claimed to have distinct inductance profiles and hence can be counted.

Advantages

This type of devices has the ability to count bicycles (vehicles) moving side by side and is highly accurate in doing so.

Disadvantages

This sensor requires shielding from weather. Presence of metal is a must and any other material can't be detected. Due to the same reason irrelevant metallic objects introduce false counts. Also distinct bicycle alloys can only be distinguished, if same material as other vehicles is used it will lead to miscounts.

Figure 4: Logger.

Products

Diamond Traffic counters and Ecocounter manufactures this type of sensors.

Pneumatic/Piezo Electric Sensors/ Tube Counters

In this method, there are two rubber tubes spaced apart. When the bicycles step on the tubes, they create air pulses which are then counted.

Figure 5: Pneumatic rubber tube sensor.

Advantages

This the most common type of sensor among the existing sensors. This is due to its portability, ease of use and high accuracy. Also these sensors are relatively low cost too.

Disadvantages

The major disadvantage with this method is the fact that rubber doesn't maintain its properties in cold conditions. The argument applies to buried pressure pads can be applied here, that whether the device can regain its original state after successive counts quickly enough. Also this sensor requires continuous monitoring and it's not vandal proof. Hence this is suitable only for short time analysis.

Products

Markman M410 and Metrocount MC 5600 are the most famous among pneumatic tube counters.

Radio Beam Sensors

Radio beam sensors are of two types: Those that can detect metal and those that are pure reflective sensors. The former can be used to detect pedestrians while the latter is used to detect pedestrians.

Advantages

Radio beam sensors are highly portable, can be easily hidden and supposedly very accurate.

Figure 6: Radio beam counter.

Disadvantages

However the accuracy is questionable under some circumstances. This type of sensors requires movement of bicycles in a single file and if bicycles were to move side-by-side false counts are possible. Furthermore as with inductive sensors, the presence of irrelevant metallic objects presents noise to this sensor.

Products

Chamber Electronics manufacture radio-beam sensors.

Infrared Sensors

Infrared beams when broken by passing bicycles can be used to count.

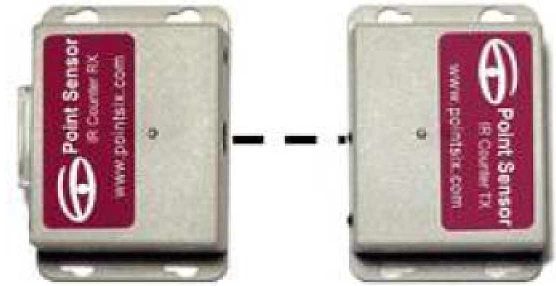

Figure 7: Infrared beam based sensors.

Advantages

These sensors are extremely low cost, small in size easy to handle and can be hidden. They need no wires, and can run for a long time on batteries.

Disadvantages

As with radio-beam reflective sensors, this sensor also requires movement in single file. Infrared sensors are also intolerant to extreme weather.

Products

Scigiene Corporation and Point Sensor Inc. are popular for this type of sensors.

Vision-Based Sensing

Vision-based bicycle counting works by separation and classification of moving objects in a video into bicycles, pedestrians and vehicles. Once classified, bicycles can be counted.

Advantages

Since this type of counting doesn't impose any requirement on how the bicycles move or on the traffic, they can be used for any type of bicycle path in urban or rural settings. They can be highly portable and easy to assemble/disassemble. Low cost and low power are also factors in favor of vision-based counting. The biggest advantage of computer vision-based counting is the verifiability of results that is not possible with other methods. As and when a bicycle or a pedestrian is detected in a video and counted, it can be archived and be verified manually in future. In fact as part of this research this was one way in which the causes of false positives and negatives could be identified. Furthermore, having a suitable visual model for bicycles and pedestrians facilitates counting in heterogeneous traffic conditions such as in urban scenarios.

Disadvantages

Cameras need to be mounted in appropriate places and may need specialty mounts. Since a camera can be spotted it's not vandal proof. Accuracy may be compromised at low light condition unless specialty cameras are used which may incur some additional costs.

Products

There are not many vision-based products around in market. SEISR LLC has been involved in development of vision-based pedestrian and traffic counting. Miovision technologies have been using similar ideas to count vehicles and draw their trajectories in freeways and intersections.

Summary

Many different methods and products for the purpose of counting pedestrians and bicycles were analyzed. Though there are still many products in the market, conceptually they are not very different from those investigated in this survey. For future work, a vision-based system tailor made to suit bicycle detection and counting will be developed. Many factors discussed in this report will be considered while developing a highly versatile system that can perform satisfactorily in wide range of operational conditions.

CHAPTER 2

COMPUTER VISION ALGORITHM FORMULATION AND SIMULATION RESULTS

Overview

A general approach is common to many of the methods that will be described in this section. Figure 8 shows a flowchart of how every frame (image) from a video sequence or live video stream, undergoes various processing steps to result in the final step. Firstly from a sequence of images, the background image can be determined. This helps to separate all moving objects in the scene. All these objects are potential candidates for bicycles. Then a certain feature extraction method is applied on each image and since these features are selected based on their representative ability, it is possible to construct a vocabulary (database) of features that represent bicycles alone. When new objects are presented to this system, the same feature extraction method is applied and features are matched against the database to see if the matching is strong. Then a classifier makes a decision as to whether the object is a bicycle or not a bicycle. This idea can be extended to pedestrians as well.

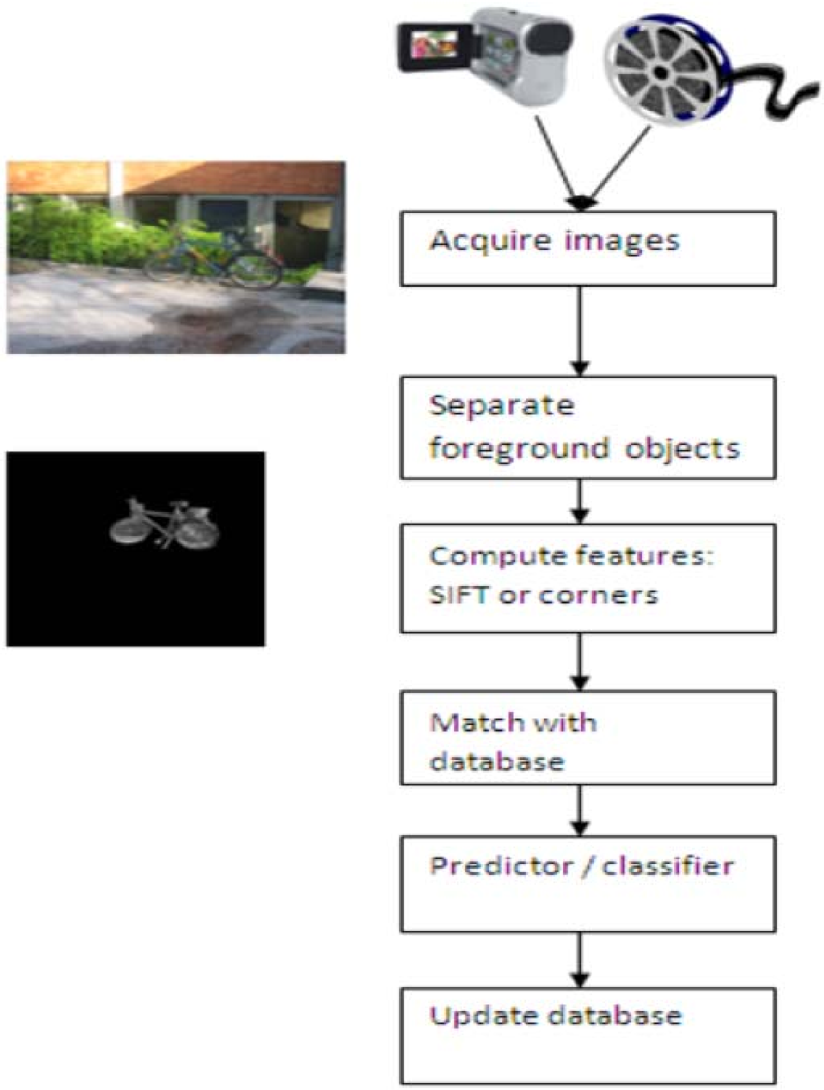

Figure 8: A flowchart of processing steps on each image/frame obtained from a video.

.

Hough Transforms

Two wheels, a handle bar and cross section are the prominent visual cues that one can think of. Hough transforms specialize in identifying known shapes. A sample result of Hough transforms for identifying circles and straight lines in a bicycle image is shown in Figure 9. The problem with this approach is the fact that circles look like ellipses in any view other than normal to the bicycle and defining thresholds for this method to consistently detect the wheels is difficult. Also this method is computationally heavy, and may hinder real-time performance.

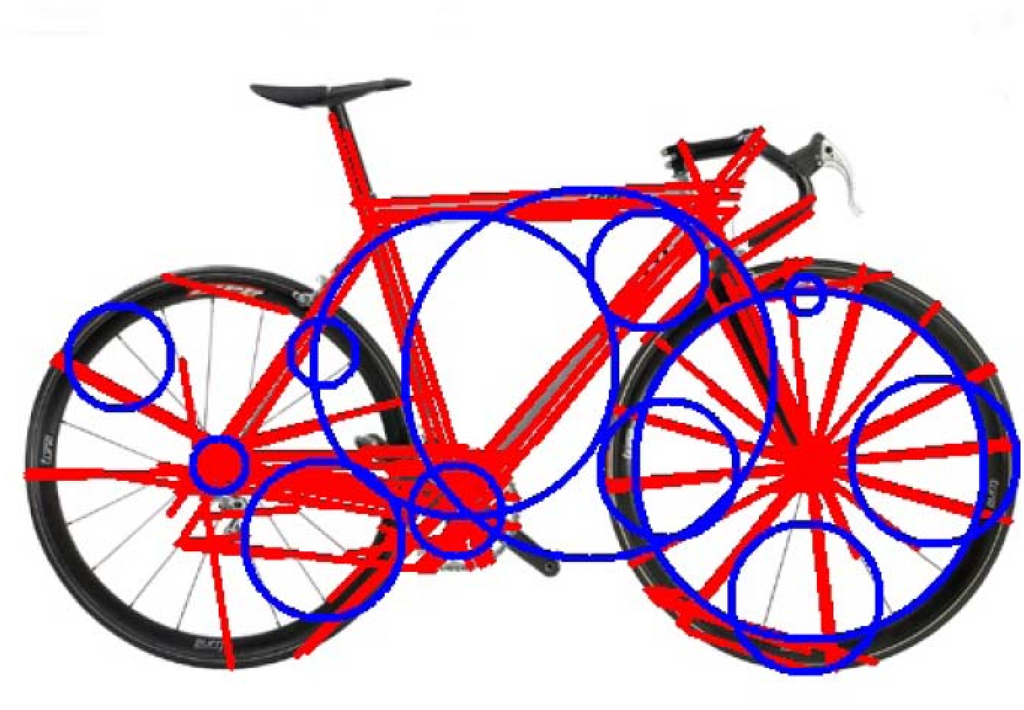

Figure 9: Hough circles and lines in a bicycle image. This image shows the problem of false detections.

Distance Transforms

Distance transforms compute the distance of each pixel to the nearest boundary pixel or feature point. This helps to construct a possibly unique template for detecting bicycles. This method is much faster to compute. However it doesn't address the problem of changes in pose. It would require explicit definition of templates for all possible views. Figure 10 shows an example.

Scale and Affine Invariant Shapes

Some recent work by Jurie et al. and Malik et al. ([2] [1]) describe shape based features and shape contexts, which are some robust regions in the image that are smaller subsections of convexities in the contours extracted from images. These regions are more immune to noise than point based features. However these methods take a lot of time to implement and test, hence they are left for future exploration

Interest Point Detectors

Scale and affine invariant interest point detectors were chosen for this problem, since they address the problem of different poses, resolutions and illumination.

Figure 10: Distance transform template of a bicycle image. Not affine invariant.

They are easier to implement and test. To establish a proof of concept, these methods were more favorable.

SIFT

SIFT stands for Scale Invariant Feature Transform [3]. This method identifies robust features and also produces a large number of features per image. The computation time for these features however is huge. The matching time for these features is also high. However for smaller images, this is a powerful method to extract features. Although this method serves our purpose, a commercial product can't be developed based on this, due to a patent on this technique. This patent involves paying a fee for a license to use this technique in a commercial implementation. There are several modifications and adaptations of this method that are also patented.

Scale and Affine Invariant Harris Corners (Haraff)

Haraff features are much less complex. They are faster to compute but require a reasonably big image [4]. Matching time is also low. And it is quick and easy to test the effect of a variety of parameters on our classification accuracy. This method can be used for vocabulary construction and matching. This method has potential for a fairly accurate real time implementation.

Experiments and Results

INRIA has a huge database of images of different classes of objects such as bicycles, people, cars etc. For the experiment 162 bicycle images and 140 people images were used. The vocabulary was constructed for the 162 bicycle images using both SIFT and Haraff features. Then 140 bicycle images and 140 people images were used for training. Some sample images from this dataset is shown in Figures 11 and 12. For training the features from these images are extracted and matched against the vocabulary to yield a histogram for each image. Then labels are given for each image manually as 1 if it is a bicycle and 0 if it is not. This process helps an SVM (support vector machine) classify bicycle images from those that aren't. Then, using a process known as cross validation, the accuracy of the classifier is determined.

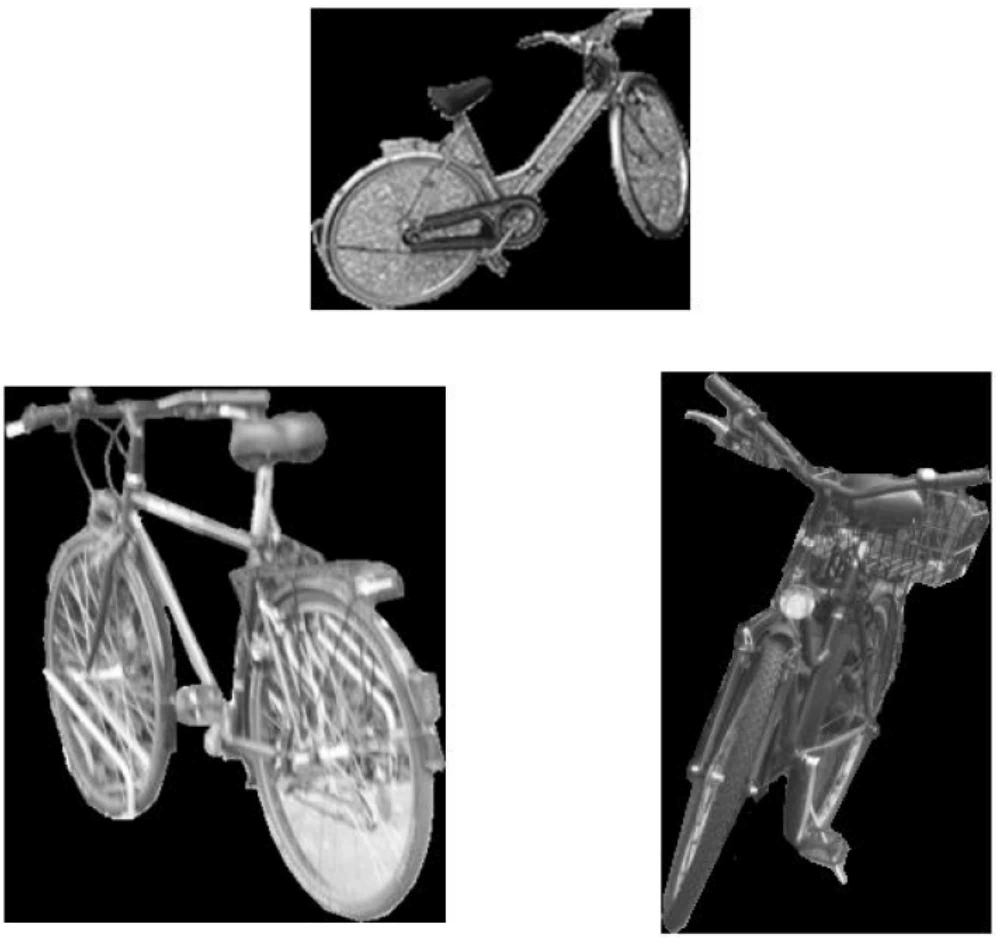

Figure 11: Bicycle images from the INRIA dataset.

Figure 12: People images from the INRIA dataset.

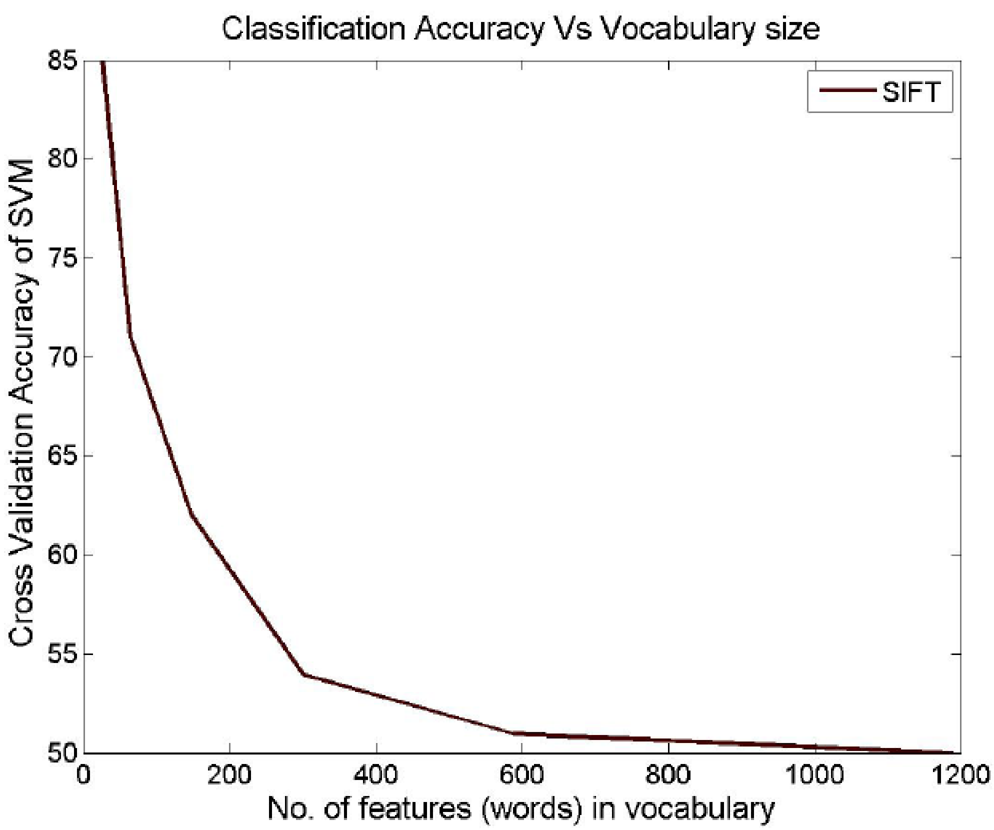

Figure 13: Cross Validation results for SIFT features based SVM.

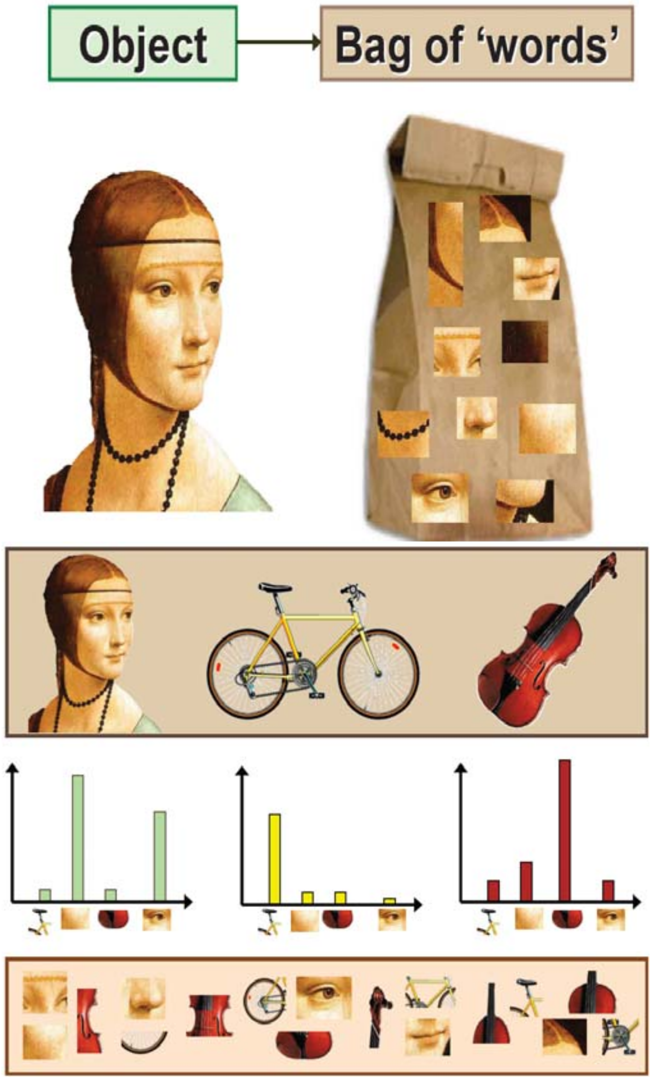

Figure 14: Bag of words model. We represent objects by composing parts and build vocabularies for different objects. The bottom most box represents the vocabulary. We can discriminatively distinguish each object class with the vocabulary.

The results of cross validation accuracy with respect to size of the vocabulary of features are shown in Figures 13 and 14. These plots explain the phenomenon of over fitting, wherein when you have too many features in the vocabulary the classifier is more ambiguous whereas with a smaller set it makes better decisions. This is a good result in terms of computation speed. And the trend is uniform for different thresholds for feature extraction. Hence designing parameters should also be quick. We can see that an accuracy of 76% is possible with Haraff features.

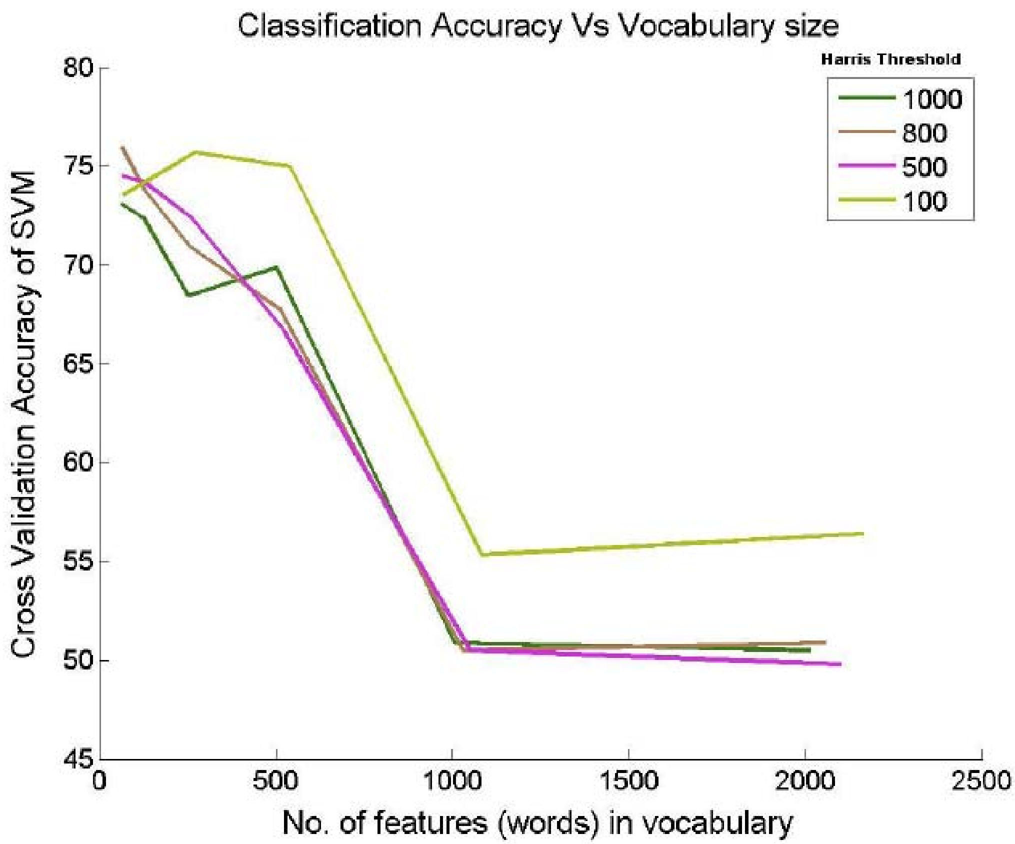

Figure 15: Cross validation results for Haraff features based SVM.

Summary

Measurements from motion such as velocity of the foreground object and the aspect ratio of the object are expected to be uniquely distributed for bicycles as opposed to other classes of objects like pedestrians and cars. Using these parameters is expected to boost the accuracy of the system. Also more videos need to be examined for consistency of results. A portable hardware system is available for multiple site data collection. Thus a generalized concept was established for detecting and counting bicycles. Then this system was tested with standardized datasets. A stand alone system that can do all the processing is currently being implemented. More data will be collected and examined using this system to test and possibly improve the accuracy.

CHAPTER 3

DATA COLLECTION AND COUNTING RESULTS

Phase 1: November

Based on the limited number of locations we could get permission to record data and time constraints of gathering some data before winter, traffic engineers and planning experts from Mn/DOT arranged for us to use the ABC Ramps in Downtown Minneapolis as the site for filming and for testing the algorithm developed so far. A total of five hours of video was collected over a period of three days. Filming was done in the early hours of around 7am to 9am chiefly, since that is the time bicyclists use the bicycle lanes near the ABC ramps to commute to work or as a transit means. Even though this is a site with frequent bicycling activity, the time of the year was not conducive for bicycling due to the weather. Filming was done on October 28th and 29th, and November 3rd. The temperatures were below freezing and resulted in limited bicyclist traffic. In fact over three days and five hours of filming, there was nearly a total of one hour of no activity at all. Further only 19 bicyclists were counted over all by manual counting. The counts generated by the software will be detailed in the next section.

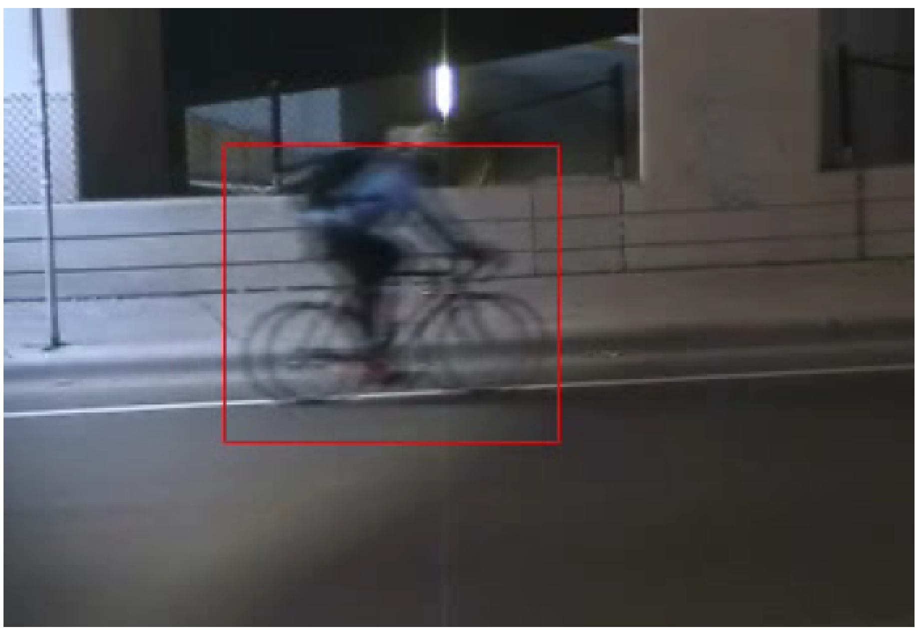

Figure 16: A snapshot of a detected bicycle using the SVM-SIFT method.

The videos from the ABC ramps site were subjected to two different methods for counting. The first method is the Support Vector Machine (SVM) which was trained using a few sample bicycle images from the INRIA dataset and some images of cars obtained from the web, since cars are the predominant opposing class of objects observed in traffic scenes such as the ABC ramps when detecting bicycles. A snapshot of a detected bicycle is shown in Figure 16. In order to compare the results of the classifier, a simplistic template matching with a bicycle template was also done. The First table shows the true counts obtained by manual counting. The results of both classification methods are tabulated in the table.

Method	Count	FP	FN	Acc.
Correlation	23	8	4	79%
SVM - SIFT	17	4	6	68%

Table 1: Count results: Phase 1.

Phase 2: Summer

Four locations were chosen for data collection. About two hours of video were collected from each site. A total of 8 hours and 30 minutes of video was collected. The locations were Pleasant St. SE at University of Minnesota (Figure 20), Bridge near East River Road (Figure 17), 15th Avenue SE and University Avenue SE (Figure 19), and finally the Greenway trail in Uptown Minneapolis (Figure 18). The sites provided a variety of training samples and situations for bicycles and pedestrians. They also provided instances of bicycles occluded by vehicles, pedestrians, other bicycles and so on. The 15th Avenue SE site had predominantly vehicle traffic. The remaining sites had mostly pedestrian and bicycle activity. Data collection was carried out using a Sony hand-held camcorder, shown in Figure 22 mounted on a tripod, shown in Figure 23. One hour mini DV tapes were used; hence tapes were changed every one hour. A good quality battery ensured data collection for up to five hours without recharging.

Figure 17: A snapshot from site: Bridge near East River Road.

Figure 18: A snapshot from site: Greenway trail.

Figure 19: A snapshot from site: 15th Avenue SE.

Figure 20: A snapshot from site: Pleasant St. SE.

Lighting and Weather Conditions during Data Acquisition

Video capture was carried out during daylight conditions in two to three hour periods between noon and 5 p.m. Hence the light conditions were quite good. As we can see in Figures 16, 19 and 20, three locations had heavy shadows. This presents a big challenge to the tracker and hence the count accuracy. But the performance was not inferior in these locations based on the observed results. In fact, these situations provide examples to build a more diversified model. We also observe that the camera angles were within a specific range. This is because; we try to maximize the resolution in the horizontal view of the bicycle. Based on our observations in simulation, for a single camera case this is the best camera angle. This view suffers from problems such as parallel riding, but this is a problem that can be overcome with multiple cameras.

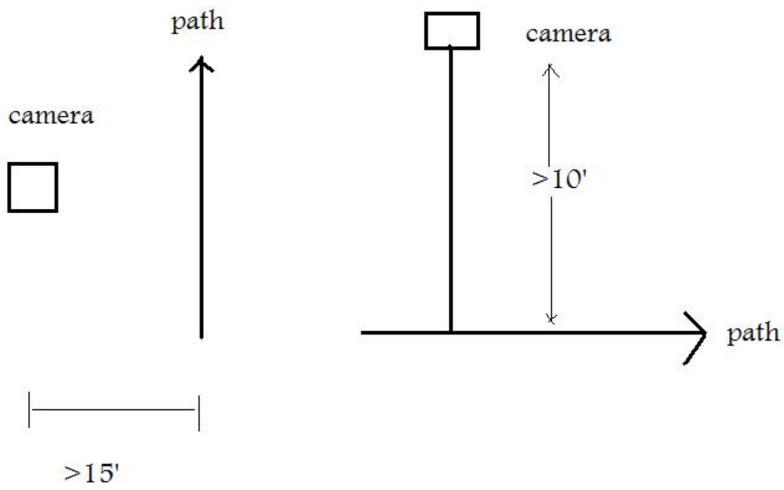

Figure 21: Typical camera placement (left: top view, right: side view).

Implementation

The implementation was carried out in C++ using open source computer vision libraries such as OpenCV and VXL. This can be employed with live camera streams or with recorded videos. All processes mentioned in the flowchart are automated with no human intervention. The training phase however required manual labeling of whether the selected blob is bicycle or a pedestrian or neither. The first step of processing is background removal and separation of foreground objects. This is done using the mixture of Gaussians method of background modeling [6]. The separated blobs are associated across different frames using a bipartite graph based approach which is based on blob area overlap percentages. Although tracking blobs is not of primary interest to this problem, it provides useful information when combined with the calibration information.

Figure 22: A Sony hand-held camcorder.

The scene was calibrated beforehand using the method described in [5]. This provides area, velocity information and other geometric information such as width to height ratio for the tracked blobs. These properties of bicycle blobs are likely to be different when compared to pedestrians. A mixture of two Gaussians is fit to the velocity and blob width to height ratio values to make a prediction on the blob's object class. This information sometimes enhances the accuracy of the classifier. At the moment this remains the only good way of resolving conflicts between bicyclists and pedestrians, introduced by the occlusion of bicycles by bicyclists themselves. Also blob tracking helps associate identities with each blob image across different frames, thus preventing duplicate counts. Blob images tend to be small. For a typical video resolution of 320 X 240, blobs are 25 X 45 pixels. Blob images are hence resized to 200 X 200 pixels.

Figure 23: A tripod.

Once the blobs have been resized SIFT features, and PHOG features are computed. The Harris corners were not produced in significant number from such small images in order to synthesize a meaningful vocabulary. Also the underlying principle of scale adapted Harris and SIFT are not very different. Hence for practical testing only SIFT and PHOG were used. These features were then utilized for prediction. Based on the type of features the processing speed was different. SIFT method was slightly slower than PHOG. With SIFT processing speed was 12 frames/second and for PHOG it was 14 frames/second. The model used for prediction can be derived from the simulation results or it can be derived from features of blobs extracted from videos with the aid of manual labeling. For practical testing 2 hours of video was captured in 3 different locations. In each of these locations 30 minutes of video was used for training. Training footage is not only used for building the appearance models for various approaches but also for extracting average blob velocity and blob width to height ratio.

The reason behind choosing width to height ratio is that for different camera placements the absolute width and height of the blob changes, but the ratio is suitable for comparison of different objects, since bicyclists are often wider than pedestrians. Despite this observation practical values seem to show that blob width, height, perimeter and area are unreliable information to classify objects in crowded scenes. This has been attributed to rapid blob merging and splitting. Also in crowded scenes, pedestrians tend to move in groups. Typical distributions of velocity and width to height ratio values are shown in Figures 24 and 25. Velocity values are hence more suitable to be used in conjunction with the appearance based classifier. These distributions are obtained from training data from 3 real traffic videos for 133 bicyclists and 146 pedestrians. Two of these scenes had bicycle and pedestrian activity only, since they were from a pathway in a university and a state bicycle and pedestrian trail. The third site was a signalized intersection which presents additional vehicle traffic. Though vehicles do not lead to misclassification due to their obvious large sizes and velocities in addition to uniform coloring, they act as partial and complete occlusions to cyclists and pedestrians in some circumstances. It is possible not to detect a bicycle or a pedestrian at all due to this effect.

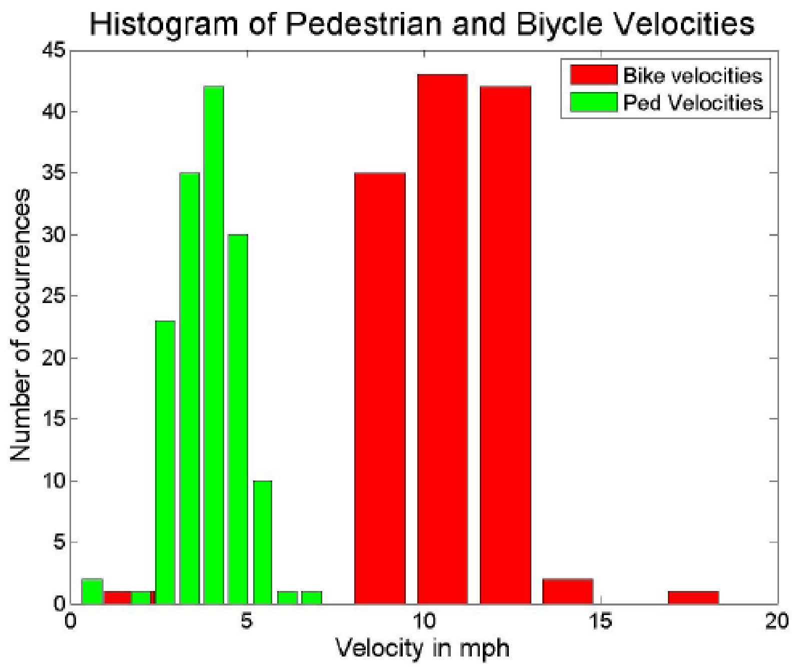

Figure 24: A distribution of velocity values for pedestrians and bicyclists in real traffic videos and bicyclists in real traffic videos.

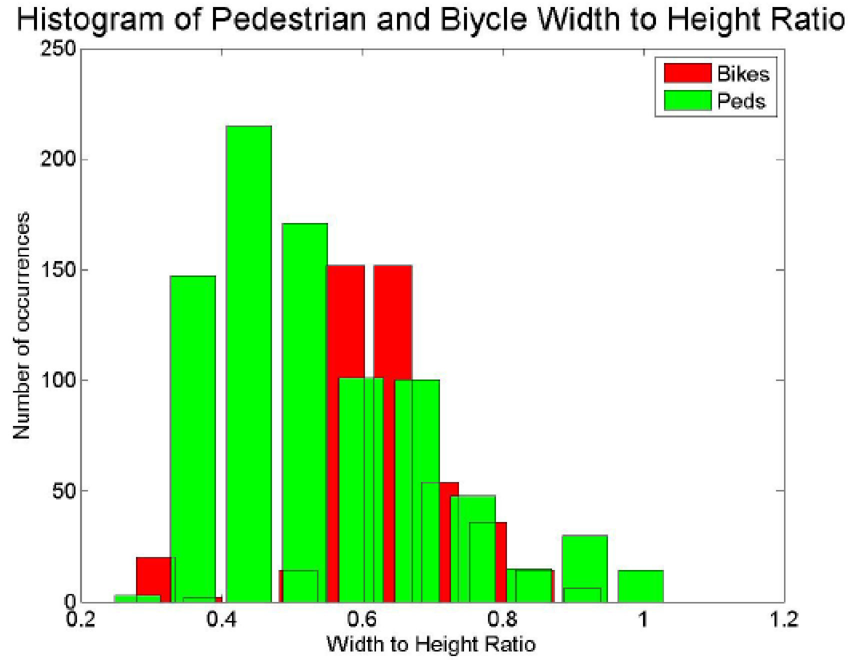

Figure 25: A distribution of blob width to height ratio values for pedestrians

Results and Discussion

More training data from real videos not only improved the visual model but also provided velocity distribution information of bicycles and pedestrians. This information sometimes improved the performance of this system. Especially when the scene was overexposed and if the bicycle falls on the shadow of a building, it appears less dense and the cyclist is more prominent leading to misclassification. The overall performance based on parts of the videos collected from (bicycle count) from the four sites (university, state trail, and street intersection) is shown in Table 2. This accounted to about two hours of overall video used for training. A total of 124 bicycles and 233 pedestrians were manually counted. The objects which were not bicycles were deemed as pedestrians. This resulted in under-counting groups of people. The accuracy with respect to pedestrians was around 70%.

Also in the state trail, a lot of bicyclists were parallel riding, reducing the count. This count was added to false negatives. The SVM-SIFT method also resulted in more false negatives than PHOG. This is attributed to the fact that a majority of the features are contributed by the cyclist in situations such as presence of shadow of a building. Velocity information was used along with the SIFT method to reduce the false negatives. This is also indicated in Table 1. Using the t rained model. Detailed testing was done on the full length videos from all sites. For this only the PHOG method has been used due to its better accuracy and also due to SIFT method being patented which in turn incurs additional licensing costs. The detailed results are shown in Table

3. We see that in the Greenway trail a lot of false negatives are produced. This is attributed to excessive parallel riding and large bicycle group riding in the trail. False positives are mainly caused due to tracking errors and blob labeling errors. Based on the observations of the false detections it can be concluded the amount of training the system has been provided with thus far is sufficient. Providing additional training can probably boost accuracy marginally, but the majority of errors are due to other factors mentioned before. Hence the primary objective of future work will be more emphasis on reducing such factors and lesser emphasis on training.

Method	Count	FP	FN	Acc.
SVM-PHOG	105	6	24	80.5%
SVM-SIFT	99	4	28	77.27%
SIFT+Velocities	107	5	21	83%

Table 2: Tabulation of results of bicycle counting with real training data

Site	True	Count	FP	FN	Acc.
University 1	66	54	4	12	75.6%
University 2	51	43	2	10	80.39%
University 3 (1 hour)	46	41	3	8	82.6%
Greenway Trail (3 hours)	495	402	21	114	76.9%

Table 3: Tabulation of results of bicycle counting: full result

The software generates a mathematical model of bicycles and pedestrians which are basically huge matrices containing distinct feature vectors corresponding to each class of object. Whenever someone carries out training this matrix is updated. If someone wishes to share the training data, they only need to exchange this matrix. The algorithm can directly work with these matrices. The program at the moment is a console application which can be executed with one simple command in the DOS prompt. This mode is only a developer mode. However adding a GUI to facilitate easy use for the average user is a straight-forward process. Due to foreseen changes to the software to deal with the problems mentioned in the discussion above, the software is still kept at development phase and has not been made available for commercial use.

CHAPTER 4

OUTREACH WORKSHOP

The first outreach workshop for this project was conducted on the 10th of June, 2009. There were close to 20 registered attendees. The purpose of this workshop was to bring forward a vision-based counting system is an attractive alternative to existing technologies for bicycle and pedestrian counting. The workshop had a presentation of the existing technologies, their pros and cons. The highlight of the workshop was a step by step description of how a practitioner can go about using a vision-based counter. (Refer Appendix B).

As a first step video acquisition was discussed. This step is very crucial for counting purposes and many different options for this step were discussed. The first being using a camcorder and tripod set up. This is a cost effective solution but is suitable for short temporary filming sessions. Also convenient locations for setting up a tripod are necessary. Other options which are more stable and suitable for longer and permanent set up were also discussed. One of them, being the portable filming system that can be transported with a medium sized car and the other being the pneumatic telescopic boom system which has a rugged build for long hours of filming and the battery to support it.

The second step was to make measurements in the scene, this aids in converting image coordinates to world coordinates, thereby aiding in speed and size calculations of moving objects. Preferably, the width of a street and known distances can help in better calibration.

The third step was to transfer the collected video to the computer de-pending on the type of filming method used. Then the video file is to be converted to a suitable format as the fourth step (Refer Appendix C for suitable codecs and formats). If a camcorder is used, for every one hour of video, it takes one hour to record the video to the computer and half an hour to convert it into a suitable format. If plenty of hard drive space is available, then video can be recorded directly in uncompressed format which is compatible with the software. The other video capture methods mentioned in the first step facilitate direct recording of the video to the computer hard drive in suitable format, hence making them faster. Using the measurements from the third step, as a fifth step calibration can be done using a GUI-based software. This software produces a calibration text file which is needed by the counting software. Lastly the counting software can be executed with the video file and the calibration file. The workshop then had details of some data collected during summer and counting results. The average accuracy of the system so far is 80%. There is still scope for improvement and the issues have been identified. The software has also been made available for Windows and Linux operating systems using the cross-platform OpenCV library. Work is in progress for more testing and conversion of the system into a deployable product.

CHAPTER 5

FUTURE WORK AND CONCLUSIONS

During the course of filming a lot of training samples for bicycles, pedestrians and vehicles were obtained. The models for each class of objects have been constructed using the methodology specified in chapter two. Currently, alternate methods for constructing these models is considered which would reduce the number of training samples required and also improve the accuracy of the classification. Research is ongoing to parallelize the study with compressed sensing theory. Also the effective combination of multiple cameras and cameras with other sensors are to be examined in order to better improve the accuracy of the system by reducing the number of false negative cited in chapter three.

Work is ongoing in order to make this system available for practical use. Firstly, we are investigating some improved algorithms which will eliminate problems such as blob overlap and parallel riding. Also we want to minimize errors we get due to tracking. If we can keep the existing setup and improve on these problems it is possible to reach 90% accuracy. Further data collection with varying camera angles and if possible with multiple cameras is also considered. The software will be made available to practitioners when the results are more appealing.

REFERENCES

[1] S. Belongie and J. Malik. “Matching with shape contexts”. *Proceedings of IEEE Workshop on Content-based Access of Image and Video Libraries*, pages 20-26, Hilton Head Island, SC, 2000.

[2] F. Jurie and C. Schmid. “Scale-invariant shape features for recognition of object categories”. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, Washington, D.C., June-July 2004.

[3] D. Lowe. “Object recognition from local scale-invariant features”. *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, Vol. 2, pages 1150-1157, Kerkyra, Corfu, Greece, 1999.

[4] K. Mikolajczyk and C. Schmid. “Scale & affine invariant interest point detectors”. *Int. J. Comput. Vision*, Vol. 60, No. 1, pages 63-86, 2004.

[5] O. Masoud and N. P. Papanikolopoulos. “Using geometric primitives to calibrate traffic scenes”. *Proceedings of the 2004 IEEE/RJS Int. Conf. on Intelligent Robots and Systems*, Vol. 2, pages 1878-1883, Sendai, Japan, Sept.-Oct. 2004.

[6] C. Stauffer and W. E. L. Grimson. “Adaptive background mixture models for real-time tracking”. *Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, June 1999, Fort Collins, CO.

APPENDIX A: TABULATION OF METHODS

Table of Methods and Properties

S No	Method	Ease of use	Bike/Pedestrian	Cost	Light conditions	Accuracy	Environmental Sensitivity	Sites	Wireless	Power	Portability
1	SEISR LLC system (Vision based)	Medium: can be serviced with palmtop or laptop	Bicycle(Vehicle)	Medium (Variable)~2500	Day/Night(with IR illumination)	Medium ~85%(*)	Weatherproof housing good dynamic range	Urban/Rural	Bluetooth	<5W	Highly Portable
2	Marksman 410 (Tube based)	Easy, Accompanying Software. Enables configuration	General Purpose (can be used for bike)	Low(around 1100 dollars)	Any	High	Not suitable for cold conditions < -5C	Urban/Rural	NA	2 Alkaline D cells	Portable
3	Metro Count MC 5600 (Tube & Electronic Piezo based)	Easy Installation	Classifier and counter of vehicles bikes including	Low(around 1100 dollars)	Any	High	Not suitable for cold conditions <-5C	Urban/Rural	NA	4 D cells for 290 days	Portable
4	EcoCounter (Hydroacoustic tube)	2-4 hours to install needs to buried in a trench pre-configurable	Bicycle	Medium Variable	Any	High	>-10C with asphalt kit waterproof	Suitable for trails	NA	Built-in 10 years	Not portable
5	Chambers Electronics (Radio Beam)detect metal RBBC4	Easy installation can be hidden	Bicycle	Medium variable	Any	Medium (Single file)	Polycarbonate waterproof housing	Any	NA	6 D cells last 45 days	Highly portable
6	Chambers Electronics (Radio beam) reflective RBBP7	Easy installation and use	Bicycle and Pedestrian	Medium Variable	Any	Medium (single file)	Weatherproof	Any	ULF to receiver connected to PC	similar	Highly portable
7	Scigiene RF infrared counter	Easy installation no wires	Bicycle/ Pedestrian	Low Variable	Any	Medium Single file	Waterproof	Any	418 MHz Transmitter	60 ft low power 1.5mW radio	Hgihly portable
8	Point Sensor IR counter Similar	Compact easy to use	Bicycle/ Pedestrian	Low Variable	Any	Medium Single file	Waterproof	Any	Radio upto 600 ft	Battery life upto 2.5 years	Hgihly portable
9	Inductive Loop profiling Counters and Accessories	Easy	Bicycle(presence of metal)	Low Variable	Any	> 95%	Needs special weatherproof housing	Any	NA	6 Weeks	Portable
10	TCS instruments mechano acoustic similar to hydro acoustic	Needs to buried Required expert handling	Bicycle	Medium Variable	Any	High	>-10C with asphalt kit waterproof	Suitable for trails	NA	Built-in 10 years	Not portable
11	EcoCounters Pyro Electric Trail sensor	Very easy to handle can be used on any ground	Bicycle/Pedestrian	Low Variable	Any	Medium	temp range -40C to 50C. Waterproof	Any	NA	Built-in > 10 years	Portable

APPENDIX B: STEPS INVOLVED IN USING THE VISION SOFTWARE

Steps Involved in Using the Vision Software

Step 1	Video Acquistion at the Site
Step 2	Make measurements such as street widths for calibration purposes
Step 3	Copy the video to a computer (depends on acquisition method)
Step 4	Convert the video into a suitable format
Step 5	Perform calibration with acquired measurements
Step 6	Run the counting software with video file and calibration file

APPENDIX C: SUPPORTED VIDEO CODECS AND SOFTWARE

Container	FourCC	Name	Description
AVI	'DIB '	RGB(A)	Uncompressed RGB, 24 or 32 bit
AVI	'I420'	RAW I420	Uncompressed YUV, 4:2:0 chroma subsampled
AVI	'IYUV'	RAW I420	identical to I420

Free converters: Mencoder, Virtual Dub