

DRAFT Whitepaper

Custom Transportation Data Collection Software for Handheld Computers

Robert L. Bertini
Department of Civil & Environmental Engineering
Portland State University
P.O. Box 751
Portland, OR 97207-0751
Phone: 503-725-4249
Fax: 503-725-5950
Email: bertini@pdx.edu

Andrew Byrd
Department of Computer Science
Portland State University
P.O. Box 751
Portland, OR 97207-0751
Phone: 503-725-4285
Fax: 503-725-5950
Email: abyrd@cs.pdx.edu

December 2003

ABSTRACT

In an effort to facilitate data collection for research and give students first-hand experience collecting data for class projects, the Portland State University Intelligent Transportation Systems (ITS) Laboratory has developed custom data collection software for PalmOS handheld computers. The software is designed to allow exporting the collected data to desktop computers in common file formats suitable for analysis in spreadsheet and geographic information systems (GIS) applications. Data collection problems addressed include recording position over time, recording the geographic location of features, and performing cumulative vehicle counts.

INTRODUCTION

The process of data collection is central to research in intelligent transportation systems (ITS), and is important both for understanding transportation phenomena and for validating the impact of transportation improvements. Data collection is also a necessary component of student research projects and it would be advantageous to include it as part of the learning process for engineers and planners. The collection of data for research tends to be costly or time-consuming, requiring many people, many hours of their time, and often expensive equipment. These restrictions lead to short durations of data collection and small data sets which negatively impact the accuracy of analysis and the resulting conclusions. The ability to collect a larger quantity of data without expending more resources would facilitate statistical analysis and improve the validity of studies.

By creating tools for transportation data collection using equipment readily available at low cost (often referred to as “commodity hardware”) and custom open-source software, students and researchers could be given access to a large pool of data-collection devices, which would give individuals or small groups the ability to collect their own data sets of statistically significant size on demand. Individual access to these devices would allow data collection to be performed on the researchers’ own schedule, even within a tight budget. Many situations (such as intersection turning counts) also require simultaneous data collection at more than one location and would be facilitated by the availability of multiple devices. Several students or researchers could also work independently to amass a large data set. Rather than depend on preexisting or general-purpose data sets, researchers could collect data specifically to suit research needs, and would have the option of collecting new data whenever it is needed.

Students would have the opportunity for firsthand experience with the data collection process when working on class assignments. [ELABORATE](#)

PROBLEMS AND OBJECTIVES

There are three broad classes of data collection problems with which we are concerned, each of which has a large number of applications in transportation research and engineering. We will discuss here only a few of these possible applications – users are encouraged to find new uses for the software and provide feedback.

The first problem class is that of recording location in one or two (or even three) dimensions over time. Applications include recording transit vehicle trajectories or analyzing changes in traffic flow using a probe vehicle. Past solutions to this type of problem have typically involved an odometer transducer or a global positioning system (GPS) receiver attached to a notebook computer which runs commercially available position logging software. The bulk

and cost of this solution does not allow it to scale well, and the software provided with most GPS receivers tends to be somewhat inflexible.

The second data collection problem is the association of features with geographic locations. If we were able to record electronically where in space different characteristics were observed and export these data in an appropriate format, the data could later be accessed with a geographic information system (GIS) either alone or in comparison with other saved data for interpretation and analysis. This is of particular interest in places like Portland, Oregon, where large amounts of transportation and planning related data are already available in GIS formats.

The third and final data collection problem we would like to address is that of recording the occurrence of individual events over time at a specific location. Some of the more common applications in this case include tabulating turns at intersections or determining the mix of vehicle types present on a facility. This general class of problems involves recording the time of each event, some descriptive category information for the event, and the cumulative number of events in this category up to the present time. These data can produce (among other things) what are referred to as cumulative count ($N(x,t)$) curves when plotted. Giving the user the ability to specify category information would allow maximum flexibility across different applications.

Our objective is the automation of these three data collection procedures by means of portable electronic devices. The data sets should be recorded on a device and in a format which allow them to be readily exported to a desktop computer for further analysis in common applications.

SOLUTIONS

We have developed two software applications which provide solutions to our three main data collection problems. All three of our data collection scenarios are location-specific, and may involve collecting data in moving vehicles. The transportation facilities and study areas are often located quite far from the lab or office. To achieve the portability that this situation demands, we decided to work with handheld computers as our hardware platform. Previous experience with notebook computers indicated that, while portable, they were too large and complex for our purposes. PalmOS-based handhelds were readily available at prices which allow the purchase of several devices, and GPS receivers are common for this platform. Open source software for PalmOS was already available which provided some core GPS functionality. The open-source development model was chosen for its compatibility with our desire to share our collection tools with the research community.

Solution 1: ITS-GPS

Our first solution, ITS-GPS, handles the first two data collection problems – recording location over time and associating features with geographic coordinates. Both of these problems depend on recording position information, so both can be solved with a GPS receiver as the main data source. ITS-GPS was built on the base of an open source GPS application called TZGPS which was originally developed by Tom Zerucha (see <http://tzgps.sourceforge.net>). Though ITS-GPS is primarily a PalmOS handheld application, the system is actually pair of programs, one for PalmOS handheld computers and one for desktop PCs. The program which runs on the handheld computer allows the user to record the presence features or events in with corresponding time and position information. The accompanying Windows application is used to retrieve the recorded data from the handheld computer, compute additional information, and save it on a

desktop machine in various file formats which may then be loaded by a spreadsheet or GIS program for display and analysis.

When the data logging software is started on the handheld computer, the user is presented with the configuration form shown in Figure 1 below.

Figure 1 - TZ-GPS Configuration Form.

At this point, the user specifies information related to the current data collection session, starting by giving the data log file a name. While recording data, the software will make an entry in this log at regular intervals specified in the field called “log delay.” By default this interval is set at three seconds. If the user wants only to record position versus time, the feature name fields may be left blank. If he or she wishes to record additional information, the appropriate feature name fields must be filled in.

There are two types of features which are handled by ITS-GPS, which we refer to as continuous and point features. Point features are the simpler of the two, and are used to record discrete events or physical objects occurring at one specific point in space. While logging data, a button will be available for each point feature specified. When a button is tapped with the stylus, the corresponding feature will be noted in the next log entry. For more accurate location of point features, the user may check the box labeled “immediate” which will force a log entry at the moment the feature is noted rather than waiting for the log delay to elapse. The program can record up to four different kinds of point features in a single data logging session. These four point features are referred to as A, B, C, and D on the configuration form.

Continuous features, on the other hand, are used to record features which persist over a distance or time (i.e. features which have length) but may change in some detail over that length. The software can record two types of continuous features in a single data logging session. Each continuous feature type is implemented as a multiple-choice selection between four different options. The types are referred to as A and B, and the choices within each type are numbered 1-4 on the configuration form. During data collection, a continuous feature is not noted at individual log entries. Instead, every log entry includes a record of whichever of the four multiple choice options is currently selected for each of the two continuous feature types.

Perhaps the best way to illustrate the difference between the two types of features is by example. If the user were interested in recording information about a highway facility, he or she might want to record the location of exit ramps, the number of lanes, and the speed limit all

along the facility. The exit ramps can be regarded as point features because they exist only at specific points, but there is a speed limit at all points along the highway. The idea is not to mark the position of a speed limit, but the value of the speed limit at every position. The same holds true for the number of lanes – it is necessary to note a number of lanes at every log entry. Point feature A could be named “exit,” while continuous feature A could represent the speed limit and continuous feature B the number of lanes. Within each continuous feature type, A1 could be “35mph,” A2 “45mph,” A3 “55mph.” and A4 “65mph.” B1 could be set to “2 lanes,” B2 to “3 lanes,” and so on.

When configuration is complete, the user can move along to data collection on the form shown in Figure 2 below.

date: 8 Dec 2003		fix: 3d	
time: 23:02:09 GMT		batt: 2.5h	
diff: +08:00		sats: 04/11	
lat: 045°00.000'		log:	
long: 122°00.000'		msgs:	
annotate		new log	
P1	P2		
35mph	45mph	55mph	65mph
2 lanes	4 lanes	6 lanes	8 lanes

Figure 2 - ITS-GPS main data logging form.

While this form is displayed, the program will be logging time and position information from the GPS receiver every few seconds. The user can note a point feature or change the currently selected continuous feature by tapping the buttons provided on the bottom half of the form with the stylus. An annotation button is provided which allows the user to associate arbitrary text information with the current position in the log file. The upper half of the display provides the user with status information while they are logging data.

Once the user returns to the office or lab, they can use an accompanying Windows application to retrieve the log files from the handheld computer, calculate distance traveled and speed between each pair of consecutive log entries, and save the log in comma separated value (CSV), dBase (DBF), or ESRI Shapefile formats on the desktop PC. CSV files are essentially a text format and can be loaded into many different applications including spreadsheet software. DBF files are intended for database software but can also be loaded into spreadsheets and geographic information systems (GIS) software. The ESRI Shapefile option allows the saved data to be imported directly into ArcGIS as a polyline representing the path taken during data recording. All recorded feature information will be associated with the appropriate line segment along the path.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	date	time	elapsed	latitude	longitude	continuous A	continuous B	none	nor	non	none	dist mi	vel mi/h	dist km	vel km/h	notes
2	2003.11.01	21:49:33	3	45.41576	-122.631	Sidewalk	no data	0	0	0	0	0.001384	1.660396	0.002227	2.672219	none
3	2003.11.01	21:49:36	3	45.41571	-122.631	Sidewalk	no data	0	0	0	0	0.003492	4.189972	0.005619	6.743286	none
4	2003.11.01	21:49:39	3	45.41566	-122.631	Sidewalk	no data	0	0	0	0	0.003753	4.503069	0.006039	7.24718	none
5	2003.11.01	21:49:42	3	45.41561	-122.631	Sidewalk	no data	0	0	0	0	0.003753	4.503069	0.006039	7.24718	none
6	2003.11.01	21:49:45	3	45.41557	-122.631	Sidewalk	no data	0	0	0	0	0.00338	4.055995	0.00544	6.527664	none
7	2003.11.01	21:49:48	3	45.41552	-122.631	Sidewalk	no data	0	0	0	0	0.003752	4.502511	0.006039	7.246282	none
8	2003.11.01	21:49:51	3	45.41548	-122.631	Sidewalk	no data	0	0	0	0	0.00338	4.055995	0.00544	6.527664	none
9	2003.11.01	21:49:54	3	45.41543	-122.631	Sidewalk	no data	0	0	0	0	0.003966	4.758978	0.006383	7.659037	none
10	2003.11.01	21:49:57	3	45.41538	-122.631	Sidewalk	no data	0	0	0	0	0.003753	4.503069	0.006039	7.24718	none
11	2003.11.01	21:50:00	3	45.41533	-122.631	Sidewalk	no data	0	0	0	0	0.003753	4.503069	0.006039	7.24718	none
12	2003.11.01	21:50:03	3	45.41529	-122.631	Sidewalk	no data	0	0	0	0	0.003127	3.752446	0.005033	6.039137	none
13	2003.11.01	21:50:06	3	45.41524	-122.631	Sidewalk	no data	0	0	0	0	0.003967	4.760033	0.006384	7.660735	none
14	2003.11.01	21:50:09	3	45.41519	-122.631	Sidewalk	no data	0	0	0	0	0.003753	4.503069	0.006039	7.24718	none
15	2003.11.01	21:50:12	3	45.41515	-122.631	Sidewalk	no data	0	0	0	0	0.003127	3.752446	0.005033	6.039137	none
16	2003.11.01	21:50:15	3	45.41511	-122.631	Sidewalk	no data	0	0	0	0	0.00338	4.055995	0.00544	6.527664	none
17	2003.11.01	21:50:18	3	45.41506	-122.631	Sidewalk	no data	0	0	0	0	0.003753	4.503069	0.006039	7.24718	none
18	2003.11.01	21:50:21	3	45.41502	-122.631	Sidewalk	no data	0	0	0	0	0.003127	3.752446	0.005033	6.039137	none
19	2003.11.01	21:50:24	3	45.41498	-122.631	Sidewalk	no data	0	0	0	0	0.003127	3.752446	0.005033	6.039137	none
20	2003.11.01	21:50:27	3	45.41496	-122.631	Nosidewalk	no data	0	0	0	0	0.001466	1.758797	0.002359	2.830585	none
21	2003.11.01	21:50:30	3	45.41495	-122.631	Nosidewalk	no data	0	0	0	0	0.000843	1.01214	0.001357	1.628925	none
22	2003.11.01	21:50:33	3	45.41491	-122.631	Nosidewalk	no data	0	0	0	0	0.002932	3.518308	0.004719	5.662319	none
23	2003.11.01	21:50:36	3	45.41487	-122.631	Nosidewalk	no data	0	0	0	0	0.002933	3.519022	0.00472	5.663467	none
24	2003.11.01	21:50:39	3	45.41483	-122.631	Nosidewalk	no data	0	0	0	0	0.002931	3.517595	0.004718	5.66117	none
25	2003.11.01	21:50:42	3	45.4148	-122.631	Nosidewalk	no data	0	0	0	0	0.00213	2.556019	0.003428	4.113623	none

Figure 3 - Example exported data from ITS-GPS viewed in Excel.

Applications Of ITS-GPS In Research And Practice

ITS-GPS has been used in many research and professional projects by students and professionals. Below we will offer some of these examples of how the software can be used.

First, we describe an example of the use of feature recording in research. Urban and Regional Planning student Mike Rose was investigating the possibility of pedestrian improvements to a road in Southeast Portland. Several segments of this road are entirely lacking sidewalks. While walking down both sides of the road, he used ITS-GPS continuous feature logging to record where sidewalks were present and absent. Figure 3 shows an example of ITS-GPS exported data viewed in Excel. These data were retrieved from the handheld computer as a shapefile and loaded into ArcGIS along with layers containing roads, tax lots, and transit stops. The total length of the missing sidewalk segments could then be calculated and an estimation given for the cost of completing the sidewalks in this area. Viewing the sidewalk information together with the transit and tax lot information also allowed an analysis of the impact of the missing sidewalks on pedestrians wishing to reach bus service or patronize one of the businesses along this road.

Figure 4 below is a map generated from the GIS software using these data. Sidewalk information can be seen along both sides of the road, with red segments indicating areas lacking sidewalks and blue dots indicating transit stops. If the transit stop information were not already available as a GIS layer, it could have been recorded with ITS-GPS point features during the same data collection session.

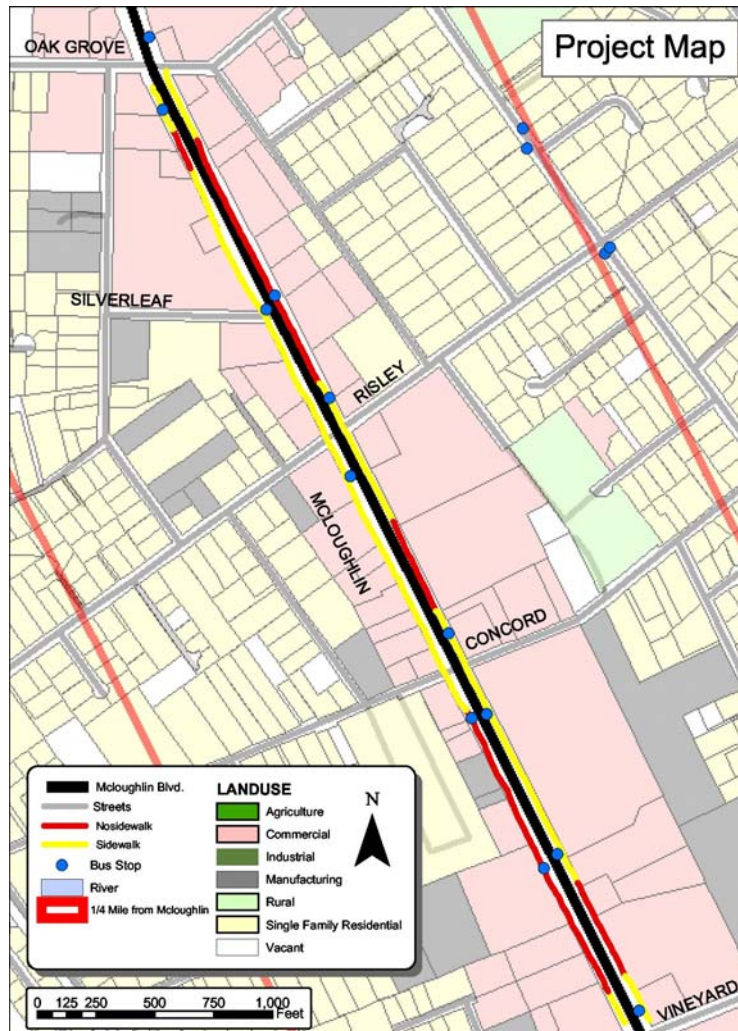


Figure 4 - Presence or absence of sidewalks along a road segment in Southeast Portland.

Portland State University PhD student Ahmed El-Geneidy has used ITS-GPS similarly for a study on factors influencing property values in Portland. This was in fact the project which prompted the continuous feature recording capabilities in ITS-GPS. One of the factors to be considered in his analysis was the presence and type of noise barriers along highways. ITS-GPS continuous features were used to record the observed noise barrier types while driving along highways in the areas to be studied.

ITS-GPS has also been used by professional engineering firms. DKS Associates' offices in Portland, Oregon used the more basic position logging functionality as part of a traffic study in Sherwood, Oregon. No continuous features were logged by the user in this particular case. The speed data calculated by the retrieval program were plotted using a color scheme in a GIS system alongside the arterial where the data were collected. The vehicle recording the data made six passes down the length of the road, and all six speed profiles were plotted together for comparison. DKS have used ITS-GPS in similar ways in Berkeley, California and for recording light rail transit vehicle speed profiles in Tacoma, Washington.



Figure 5 - Speed data recorded during multiple vehicle runs in as part of a study in Sherwood, Oregon.

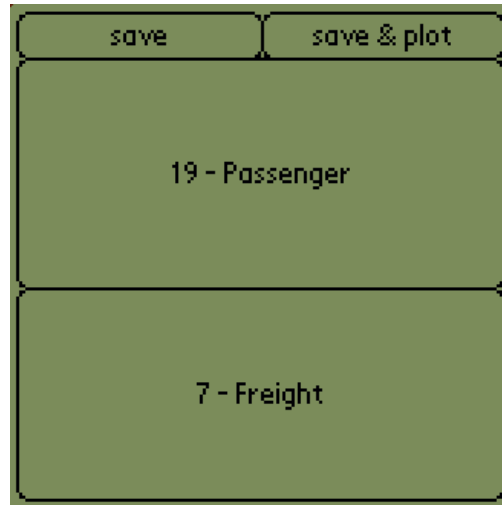
Solution 2: ITS-Count

Our second data collection solution, entitled ITS-Count, deals with the problem not addressed in ITS-GPS – the counting of different classes of events over time. ITS-Count is similar to ITS-GPS in that it consists of both a data collection application for the handheld computer and a data retrieval Windows application for the PC.

A screenshot of the ITS-Count configuration form. The form is titled "its-count: configuration" and contains several fields and buttons. At the top is a button labeled "manage saved data". Below that is a field for "new file name:" with the text "Traffic-l-84" entered. Underneath is a field for "category names for this file:" with two entries: "1 Passenger" and "2 Freight". Below that is a field for "snd volume:" with a row of five buttons labeled "0", "1", "2", "3", and "4". At the bottom is a large button labeled "start counting".

Figure 6 - ITS-Count configuration form.

Before data collection begins, the user is presented with a configuration form (see Figure 6 above) where he or she may set options related to the current data collection session. The “manage saved data” button at the top of the configuration form is used to plot, view, and delete previously saved data sets. A descriptive filename is entered for the data about to be collected, and the event categories are described on the following two lines. In the above figure, the arrivals of passenger vehicles and freight vehicles on a road are being counted separately, so the categories have been named accordingly. The user need not provide names for both categories if he or she will be recording only one type of event. Once the configuration is complete, the user may move on to collecting data using the form shown in Figure 7 below.



save	save & plot
19 - Passenger	
7 - Freight	

Figure 7 - ITS-Count in use: the event counting form.

The user may at this point either tap the buttons on the screen with the stylus or use the hardware buttons below the screen to record the occurrence of events. We found the on-screen buttons to be less than optimum as input devices when the user is not looking at the Palm display, so the four hardware buttons at the bottom of the handheld computer (which would normally call up built-in applications such as the calendar or notepad) have been reassigned to provide another method of input. Each time a button is pressed, a new line is added to the data set to record the event. This line includes a timestamp, the event category, and the cumulative count for that category. As each event is recorded, the display is updated to indicate the new cumulative count. When data collection is complete, the user can save the data and optionally plot it on the handheld device to make initial observations and provide a validity check while in the field. A sample plot is shown in Figure 8.

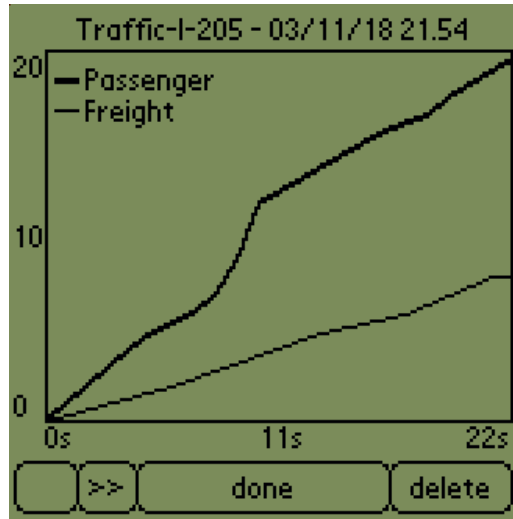


Figure 8 - Plotting a saved data set in ITS-Count.

The user can plot any saved data set currently stored on the handheld computer. The same interface used for plotting the data allows the user to browse through the saved data sets and delete any unwanted ones. Any data set may also be exported for analysis on a more powerful desktop computer, perhaps using a spreadsheet application. The Windows program included with ITS-Count will retrieve data sets from the handheld computer and save them as comma-separated value (CSV) text files. These can be readily loaded into many different programs including Excel. Figure 9 below contains an example of the exported data as viewed in Excel.

	A	B	C	D	E	F
1	elapsedTime	category	count			
2	256	2	1			
3	275	2	2			
4	277	2	3			
5	279	2	4			
6	288	2	5			
7	291	2	6			
8	302	2	7			
9	314	2	8			
10	317	2	9			
11	332	2	10			
12	336	2	11			
13	343	2	12			
14	346	2	13			
15	371	2	14			
16	374	2	15			
17	399	2	16			
18	407	2	17			
19	423	2	18			
20	427	2	19			
21	441	2	20			
22	453	1	1			
23	488	2	21			
24	529	2	22			
25	535	2	23			
26	540	2	24			
27	545	1	2			
28	568	1	3			
29	576	2	25			

Figure 9 - ITS-Count: Exported data as seen in Excel.

Applications of ITS-Count In Research And The Learning Environment

The types of count data that can be collected with ITS-Count have many different uses in transportation engineering and research. These include measuring saturation flow at intersections, performing turning counts, and finding gap times for pedestrian crossings. Manual counts from closed circuit television (CCTV) surveillance video with ITS-Count can be used to verify automatic vehicle counts from inductive loop detector systems.

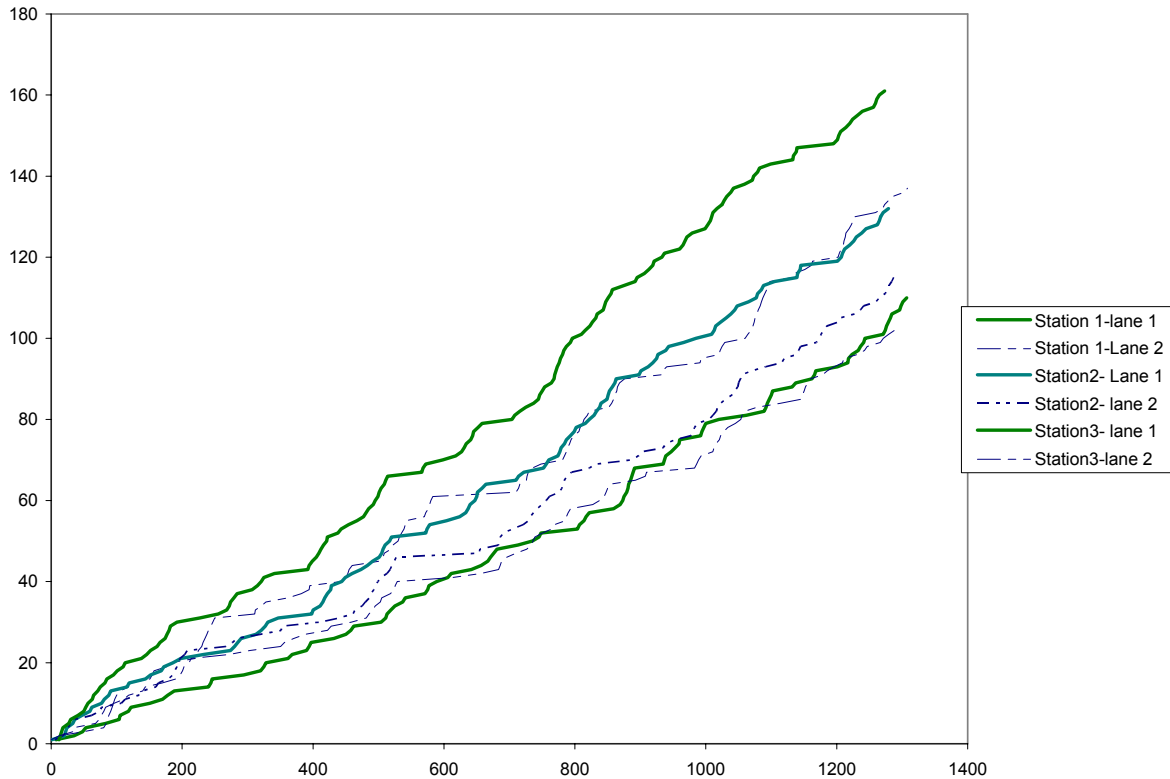


Figure 10 - ITS-Count: Plot produced by students from several exported data sets.

One of the primary purposes of ITS-Count is to allow students in transportation courses hands-on experience in data collection. Six students simultaneously recorded vehicle count data at different points along a facility and produced the $N(x,t)$ (cumulative count versus time) plot shown in Figure 10 above as part of a homework assignment. Using ITS-Count allowed them to easily record and combine their own data, making the assignment more valuable than a simple exercise with preexisting data.

CONCLUSIONS AND RECOMMENDATIONS

This project has resulted in the creation of two software packages for PalmOS handheld computers which facilitate data collection in transportation research. Use of the software has been successfully incorporated in research, professional engineering, and classroom projects.

This software has been used with the Palm V and IBM WorkPad c3 devices and the Magellan GPS Companion GPS receiver. It is likely that the software will work with other devices, but other configurations have not yet been extensively tested.

Future revisions of the ITS-Count software should include at least four event categories, as this will significantly increase the usefulness of the software.

ACKNOWLEDGMENTS

The authors acknowledge Ahmed El-Geneidy and Mike Rose of Portland State University for their input and assistance with software development and data collection. Comments from Dustin Luther of DKS Associates and Peter Koonce of Kittelson & Associates are also appreciated. This work was partially funded by the Engineering Technology Industry Council (ETIC), the PSU Department of Civil & Environmental Engineering and the National Science Foundation. [OTHERS?](#)

TABLES AND FIGURES

Figure 1 - TZ-GPS Configuration Form. 4

Figure 2 - ITS-GPS main data logging form. 5

Figure 3 - Example exported data from ITS-GPS viewed in Excel..... 6

Figure 4 - Presence or absence of sidewalks along a road segment in Southeast Portland. 7

Figure 5 - Speed data recorded during multiple vehicle runs in as part of a study in
 Sherwood, Oregon. 8

Figure 6 - ITS-Count configuration form. 8

Figure 7 - ITS-Count in use: the event counting form..... 9

Figure 8 - Plotting a saved data set in ITS-Count. 10

Figure 9 - ITS-Count: Exported data as seen in Excel..... 10

Figure 10 - ITS-Count: Plot produced by students from several exported data sets. 11