Management and mathematical models in German public transport

Günter Bärwolff Hartmut Schwandt Jan Sablatnig Tatjana Ebers Denis Klauser
Institut für Mathematik, Technische Universität Berlin, Germany
emails: {baerwolf, schwandt, sablatnig, tebers, klauser}@math.tu-berlin.de

August 6, 2015

Abstract

In the current paper we will list and discuss some mathematical problems occurring in the context of public transportation.

There are some aspects we want to consider. The first one is the problem to give a reliable documentation of the performance of a public transport company compared to others. This is the important base to satisfy the quantum of money which the companies will obtain from the government or the municipality where the companies act. As a measure for the performance of the companies serves for example the sum of number of people-kilometres.

The second aspect is that for safety and security matters, it is advantageous to gather information about the actual filling of vehicles to have a base for reaction in case of an emergency.

A third topic consists in an efficient conducting of track vehicles by the development of optimal strategies for the driver depending on time tables, stop signals, random disturbances and the actual filling of trains.

Another point is the supply of maximal information for passengers who want to use means of transportation for example to find seats or want to change for another line.

The results of our past and future work we work out together with mid size companies which act in the field of public transport.

Index Terms—People counting, online filling analysis, optimal strategies for track vehicles

1 Introduction

Automatic passenger counting with high precision raw-data is the ideal base for accounting and planning transport performance. Today, authorities and transport companies demand solutions for precise, comprehensive and automatic passenger counting compared to random sampling methods used over the past years. Automatic passenger counting provides a complete and accurate view of the transport performance and revenue distribution.

There are currently two major sensor principles in commercial use for 3D-image-based automatic people counting:
i) Infra-red technology based on the Time-Of-Flight (TOF) measuring technology.
ii) Video-technology to get stereo information as a base for tracking and counting.

Both technologies yield depth and intensity images of the monitored area. Counting algorithms then attempt to use the depth map to count individual humans that board or alight the vehicle. This yields an integer number of persons for each direction.

The commercially available sensors have built-in counting algorithms and can thus be used to only gather counting data without ever accessing the actual image data of the sensor such as the depth map. The generated people counts are then gathered and added up for all doors of the vehicle. The resulting numbers are then assumed to coincide close enough with the actual number of people on board of the vehicle and derived data can then be calculated such as the number of passenger kilometers served during the entire ride.

The actual accuracy of the counting system is very hard to determine except by extremely expensive human validation procedures. Therefore, the actual accuracy is usually disregarded and instead the symmetry is used, as this is easy to determine from the gathered data:

$$E_{\text{sym}} \quad := \quad \frac{c_{\text{big}}}{c_{\text{small}}} - 1 \qquad (1)$$

Where c_{big} is the bigger one of the counts for boarding and alighting passengers, while c_{small} is the smaller one. The metric is usually specified as a percentage. The motivation for this metric is that if the vehicle starts and ends empty, then the total number of passengers entering the vehicle must, at the end of the trip, match the total number of passengers leaving the vehicle. The number is usually given as a percentage. It is important to note that because the counts for the entire trip are added before this calculation is performed, this symmetry is usually much higher than the actual accuracy of an individual count event. It is also obvious that several trivial approaches (e.g. always counting zero, or only counting boarding passengers and reporting that same number for both directions) would yield 100 % symmetry, yet not represent a good passenger count.

Currently, the symmetry of the sensors' builtin counts is usually about 95 %. The purchasers for such technologies i.e. transportation/railways companies or the municipalities are increasing accuracy requirements and are now asking for symmetry values of of 99 % or better.

To reach values this high, some basic approaches in passenger counting will have to be revised. We are currently investigating the required advances.

We are focussing on the infra-red technology (TOF, see also [**?**]) and by the additional use of important Meta-data beyond the sensor we were successful to improve the quality of counting results already. We do this in a close cooperation with two companies which are very active in the field of hard- and software development for the configuration of public transport vehicles with counting and evaluation devices.

2 What currently counts

Traditional counting algorithms yield only integer numbers of people boarding or alighting from the vehicle. This is also the case for the built-in counting algorithms of commercially available sensors. They are generated using the depth or intensity data of the TOF sensor or the stereo information of the video sensor respectively.

3 More than counting

First of all it is important to realize that no automated system (and no human observer, for that matter) can reach a 100 % count accuracy. There are many unclear situations in which the actual number of passenger transitions can only be guessed. The reasons can be based on the hardware measuring process, e.g. random noise in the data stream, short-term overload, or presence of a objects that the specific sensor technology cannot handle well, such as mirrors for stereoscopic technologies and IR-absorbant clothing for TOF technologies. Or they can be based on the counting algorithms shortcomings, such as inability to discern between two humans moving close together and a human and a large backpack, or the inability to follow the complex movement within a crowd. But the unavailability of perfect counts can also be a direct result of the counting definition, e.g. a child might be counted differently when entering in a stroller than when alighting on foot. It is by the way also clear that this is not an error of the counting definition as the only definition that is free of cases of doubt is the definition to count every single human including all babies (which cannot be counted well at all because of their small size and because they can be inside of a buggy).

However, the realization that measurement processes of all kinds usually include some uncertainty is by no means new and the means to handle this situation are well known.

To begin with, one has to realize that a counting algorithm does not yield a count, it yields a count *estimate*. This is a guess, based on the avaiable data, what the *average* count is (with the average taken across all situations).

As an example, let us assume that simply by missing some data (for geometrical reasons, the event takes place near the border of the counting area), there are ten different counting situations that all look the exact same on the captured sensor video stream. Let us further assume that in 3 of the situations, a passenger entered the train and in 7 situations, no passenger entered the train. The real count estimate for this sensor video stream should then be 0.3 .

Traditional counting algorithms would (if well-tuned) report zero if they were tasked with such the sensor stream in question, because that is the closest integer number to the real estimate. What this does end up doing is adding a integer noise of ± 0.5 to all count data. If the number of events for the tallying of a train ride is sufficiently large, then the law of large numbers indicates that the symmetry metric is not changed by this noise. For smaller numbers of individual count events, the added noise does however add statistical noise to the symmetry measurement, meaning the measurement gets worse.

In our algorithms we allow for counting results which are rational numbers. Thus we do not round up or round off in fuzzy situations. Based on experiences and the analysis of adequate samples such non entire counting results are generated by our algorithms. It should be noted that this increases the complexity of the code, as cases that were hitherto simply pruned now have to be calculated completely. For the all-day final result of the transportation company performance we have to average and extrapolate counting data and the type of numbers is not an issue. The online use of data is also not restricted to natural numbers.

The general scientific procedure would advise to also add a confidence interval to the measurement (or counting estimate, in this case). We are investigating this for reasons outlined in 4.

Calculating a confidence based on the decision process in the algorithm is by no means an obvious procedure and therefore we opted to *measure* the confidence. For this, we are providing a large set of data gathered while the counting algorithm calculates its count estimate. This set contains data such as the SNR between the original (incoming) image and the prefiltered image, the amount of pixels changed during morphological operations, the amount of tracked objects, and the number of possible object joins. The counting algorithm then counts a large amount of situations and the differences between the algorithm's count estimate and the actual count are collected alongside the internal data set for each case. Using classification of data set cases and factor analysis it is then possible to estimate the correctness probability from the internal data set.

4 Strength Through Diversity

We used two very different algorithms to count the same set of data streams and compared their results and especially their counting errors.

A Pearson correlation index revealed a *negative* correlation betwwen the two counting errors. This indicates that a better result can be reached by combining the two counting results. Consequently, a simple arithmetic average of the two counting results already yielded a better counting result overall.

We developed this idea into a more general approach that is similar to expert system. In general, we use N counting algorithms, each producing a count estimate c_i. These results need to be combined to result in a single count estimate c. We base this combination on the normal scientific method of combining measurements (see for instance [**?**]).

Assume that several measurements exist for the same value, each represented by $x_i \pm u_i$, where u_i is the uncertainty, implying that 95% of all such measurements do include the true value in the interval $[x_i - u_i, x_i + u_i]$. Then to combine the measurements:

$$
\begin{aligned}
s_i &:= u_i^{-2} && (2)\\
x &= \sum s_i x_i / \sum s_i && (3)\\
u_{\text{int}} &= \left(\sum s_i\right)^{-\frac{1}{2}} && (4)\\
u_{\text{ext}} &= \sqrt{\frac{\sum s_i (x_i - x)^2}{(m-1)\sum g_i}} && (5)\\
u &= \max(u_{\text{int}}, u_{\text{ext}}) && (6)
\end{aligned}
$$

Where s_i is a significance, x is the weighted average, u_{int} is the internal uncertainty, u_{ext} the external uncertainty and u the combined uncertainty. Note that normally, u_{int} dominates the uncertainty, with large values of u_{ext} indicating a discrepency between the measurements going into the combination procedure. With 95% probability, the resulting measurement of $x \pm u$ covers the true value measured.

The major challenge when applying this procedure is that each measurement c_i needs to include an uncertainty, estimating which requires some additional work. The advantage is obvious of course. In those cases where a counting algorithm usually counts very well, that algorithm's results are preferred over those algorithms that are not certain in this case. This is similar to the standard ai approaches of generating many possible answers in parallel and then ranking the results to find the most likely answer.

The method described works best if the individual counting algorithms work differently, rather than similar. For instance, if all algorithms use the same filtering as an initial step, then an object that gets completely removed by that filtering is not visible to any algorithm and therefore there the combined system may not have a hint that something is wrong. If even one algorithms works with a completely different filter, chances are that that algorithm yields a different, discrepant count estimate and the overall system can at least indicate that there may be a problem by yielding an increased uncertainty (caused by increased external uncertainty).

Because it is therefore important that the different algorithms use different basic ideas to produce a count estimate, we shall explain the basic ideas of some of the main algorithms we use or are developing for this purpose.

5 Volume Based Counting

The volume based counting is realized by an algorithm that uses the distance data obtained from Time-Of-Flight sensor for searching the most possible height segmentation level of an object or person. The first segmentation level is on the 1 meter over the bottom and the last segmentation level is on the highest detectable position of an object. After the segmentation in each height level connected component pixels will be formed as pixel blobs. Using the method an adaptive segmentation threshold for an detectable object will be obtained. The last or highest pixel blob represents distinguishing feature of an segmented object and so will be used for the object tracking and counting. The resolution of the height segmentation levels plays an important part for an correct extraction of the connected component pixels as distinguishing feature. Finally an optimal resolution of the height segmentation levels reduces incorrect object counting.

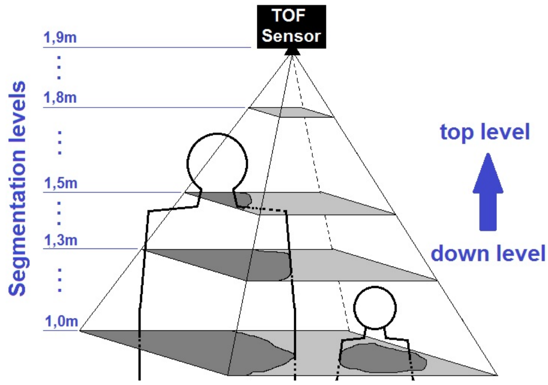

Figure 1: Segmentation threshold levels

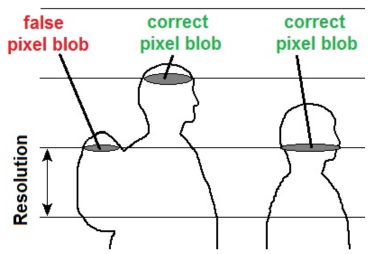

Figure 2: Resolution of the segmentation levels

6 Surface Based Counting

7 Image Based Counting

8 Metrics

As mentioned, the metric in (1) is used because it is easy to take, but it is too weak to optimize a counting system. Instead, a simple absolute error metric should be employed:

$$E^{\mathrm{b,a}}_{\mathrm{abs}} \quad := \quad |c^{\mathrm{b,a}}_{\mathrm{est}} - c^{\mathrm{b,a}}_{\mathrm{cor}}| \tag{7}$$

Here, $c^{\mathrm{b,a}}_{\mathrm{est}}$ is the estimated count for boarding or alighting passengers, while $c^{\mathrm{b,a}}_{\mathrm{cor}}$ is the correct count.

In many cases, a relative error metric is also useful:

$$E_{\mathrm{rel}}^{\mathrm{b,a}} := |\frac{E_{\mathrm{rel}}^{\mathrm{b,a}}}{c_{\mathrm{cor}}^{\mathrm{b,a}}} - 1| \tag{8}$$

This metric is usually given in percent.

There are two problems with the metrics defined. First of all, the numbers $c_{\mathrm{cor}}^{\mathrm{b,a}}$ are hard to find, especially when the algorithms are meant to be trained for large amounts of counting situations. This is one of the major cost factors for the development of improved counting algorithms.

The second problem applies to the graining of the data. The numbers $c_{\mathrm{est}}^{\mathrm{b,a}}$ and $c_{\mathrm{cor}}^{\mathrm{b,a}}$ all contain *several* count events. In their currently most fine-grained version, the numbers and hence the error would apply to individual door-openings. I.e. One door opens at a station and ${}_{\text{door open}}c_{\mathrm{cor}}^{\mathrm{b}}$ would be the amount of people boarding the vehicle through that one door at that stop while the door is open. A courser graining would be to define the counts as to apply to a station, i.e. to sum all the individual door openings at a single station:

$${}_{\text{station}}c^{\mathrm{b,a}} := \sum {}_{\text{door open}}c_i^{\mathrm{b,a}} \tag{9}$$

Even courser metrics can be defined by applying the numbers for the entire *trip* of the vehicle or even the total numbers for an entire *fleet* over a certain amount of time like a *day* or an entire *year*.

It is obvious that the courser the metric, the meaningful it becomes in assessing the counting quality. In particular, many of the older counting systems, while much less precise than the current counting systems, still manage to reach comparable errors at a sufficiently course level. The reason is that even if each counting operation has a broad statistical scatter, the law of large numbers ensures that as many operations are summed up, the estimate still gets relatively close to the correct numbers as long as the scatter is not statistically biased.

One problem in particular with using course grained metrics is that for most applications, the relevant quantity is *not* the actual number of boarding and alighting passengers but rather derived quantities such as the current number of passengers on the train or the passenger miles, i.e. the sum of distances traveled times the passenger numbers. Depending on the distribution of the statistical scatter of the counting system, the passenger miles may *not* be close to the correct number of passenger kilometers, even though the number of entry and exit events are close to the correct values.

As an extreme example, it would be possible to only install *one* sensor in a three-doored train, multiply the measurement by three and use that as the per-station measurement. While the numbers reached this way would typically be wrong at every single station, the total number over the course of the day would still be within a few % of the correct number, as long as no special bias is active such as there exist more stations where people typically enter via the front door than there are stations where people typically enter via the rear door. So the one-sensor approach would be appropriate if only the total numbers of passenger entries and exits per day or even year were needed. However, the method would typically give unacceptably bad estimates of the number of passengers currently inside the vehicle. If, for instance, the very first station of the trip has people typically enter the vehicle by the rear door, which isn't counted, then the measured fill level of the vehicle is significantly lower than the correct fill level. On one of the later stations, passengers only enter the train via the front door, thus overestimating the entry numbers and evening out the error for the entry numbers introduced at the first station. Since the train reaches its terminus soon after that, however, the passenger mile numbers are permanently off.

It is therefore advisable to choose the finest available graining, which is usually a "door open" graining. The disadvantage of using this graining is that the remaining graining depends on the specific journey taken by the vehicle. If the specific trip has a lot of long door openings, then the graining is coarser than for a trip with very short door openings. Thus, the measuring system's error would be hidden better on the first trip. This implies that the results of this measurement are not comparable to any other measurement unless there is an exact definition of the trip.

This is not a problem if one simply wishes to train a counting algorithm via a data bank with sensor streams since the trip is well-defined, it is the trip taken while the data was recorded. On the other hand, "door open" measurements retain the problem of keeping some graining, which still hides some of the counting system's error, making it unnecessarily difficult to find problems of the counting system.

The ideal metric would therefore be a grain-less system, where the counting system's performance is judged for each passenger event individually. This is currently not possible because there is no way to identify which of the passengers a counting system just counted when it sent off an event.

9 Identification of Count Events

There are two motivations to attempt to identify the count events from different counting algorithms with each other. One was given above and has to do with providing a better, finer metric. The other is that for the count combination method described in 4, it would also be advantageous to do this combination per individual count event. This is because the confidence interval is different for every single count event during a door opening. If the combination is performed on the basis of a full door opening, one first has to calculate the confidence of the full stream. This confidence will be dominated by the worst confidence in the stream. This procedure loses a lot of information and the resulting confidences are going to be quite low. As an example, a stream with two passengers boarding the train could be counted by two different algorithms. The first algorithm would see the first passenger very well, but be very uncertain of the crowd entering the train later. On the other hand, the second algorithm is the other way around, it counts the second passenger perfectly well and with high confidence. If the two measurements are added on a "door opening" basis, the confidence of the final result would be very low and the combination algorithm would have no way to know that it should choose the count results of the first algorithm for the first situation and the counting results of the second algorithm for the second situation.

We are therefore investigating methods of splitting up the measurement result so that individual, well-defined count events are generated. The obvious candidate to perform this split is *time*. Unfortunately, different counting algorithms send off the count event at different times. This is because different algorithms use different definitions about when to send off an event. To avoid this problem, one should adjust the counting algorithms so that the counting event is generated for the exact time that the passenger cross the door-line.

Unfortunately, there is some uncertainty to that time, so several counting algorithms, all using the same definition, would still not return the exact same timings for the same event. The identification therefore needs to have some temporal range within which several events are assumed to be the same.

This leaves the problem of several counting events at nearly the same time, which could then be ordered differently by different counting algorithms. To discern these events, we add the *place* where the counted object crossed the door-line. Several counting events in rapid succession necessarily occur at different places because two objects cannot be in the same place.

10 Online data for optimal driving strategies

Beside the evaluation of the transportation performance the railway companies are interested in running their business economically. That means it is interesting to save Diesel or in general energy depending on the actual filling or load of the train. To realize this the driver or an autopilot need online information about allocation of the train by people.

We will give here a brief overview on the mathematical model of train movement and the optimization goal. To illustrate the basic principle we write down the model for a simple track situation (see [**?**]). The train movement can be described by the following system of differential equations:

$$\frac{dx}{dt} = v \qquad (10)$$

$$\frac{dv}{dt} = u_B u - f_w(v) - i_m g \qquad (11)$$

where $x(t)$ and $v(t)$ are the position and velocity of the train at time t as state variables. u is the effort K applied by the train at the traction wheels in relation to the absolute value of the maximum braking effort of the train K_B

$$u = \frac{K}{K_B}$$

and serves as a control variable. $u_B = \frac{K_B}{m}$ is the specific maximum braking rate, and m the train mass. f_w is the specific train running resistance and should be fitted by the formula

$$f_w(v) = \alpha v^2 + \beta v + \gamma$$

with fitting parameters α, β, γ. i_m is the gradient of a considered section and g is the gravitational acceleration.

It is important to note that the system (10),(11) with appropriate initial data for x and v a non linear initial value problem which has no closed-form solution. The non-linearity and the possible time dependence of the section gradient and the train mass are reasons for that.

If we don't simplify the system (10),(11) by a linearisation which is proposed in [**?**] and we can't assume constant parameters m, i_m the system must be solved numerically.

The optimisation criterion shall be the mechanical energy needed to drive the vehicle. It can be computed from the state variable v (speed) and the control variable u as follows

$$\begin{aligned} Q(u,v) &= \int_0^{t_1} \frac{1}{2}(K + |K|)v\,dt \\ &= \int_0^{t_1} \frac{1}{2}(u + |u|)u_B m v\,dt\ , \end{aligned}$$

where t_0 is the starting time for a certain regime and t_1 is the switching time to change the driving regime. Now the optimisation problem reads

$$\min_{u \in U} Q(u,v) \qquad \text{subject to} \quad \text{system} \ \ (10),(11)\ , \tag{12}$$

where U is the set of admissible controls. The solution of (12) using Pontryagins maximum principle is only possible in a closed-form with very strong simplifications of the model. For a real world driving assistance system based on an adequate mathematical model for the state variables x and v we propose efficient solution strategies developed in our institute.

Adequateness means in this context the time or section dependence of the train mass which we get by the online registration of drop-outs and drop-ins.

11 Further Discussion

In the current paper we reported about our activities in online registration of passengers of public transportation systems. The improvement compared to the state of the art consists in the usage of Meta data beyond the people counting sensors to increase the accuracy to the required percentage.

Otherwise we combine information of people counting with the driver assistance systems to run trains by optimal strategies.

12 Acknowledgement

The authors gratefully acknowledge the support of Federal Ministry for Economic Affairs and Energy of Germany and federal state of Berlin/Investitionsbank Berlin for the project based cooperation with Interautomation Deutschland (IAD GmbH) and Infrared Intelligent Sensors (iris GmbH).