# High-velocity optical flow

Joris Vergeest
Delft University of Technology
Landbergstraat 15
Delft, The Netherlands
j.s.m.vergeest@tudelft.nl

## ABSTRACT

Optical flow is widely used to estimate the velocity of objects relative to a digital camera. Most commonly, two images taken with the same camera at small time difference are compared in order to detect the displacement of structures in 2D image space. Such displacement could be a measure of displacement, or motion, of objects in the scene relative to the camera. At high velocities, the displacement in image space is relatively large and the correlation of image structures gets more difficult. The displacement can be reduced by reducing the time difference, or increasing the number of frames taken per second. However, due to the reduced exposure time, the quality of the individual images gets poorer. In some practical situations, it appears technically very difficult to achieve reliable speed measurement at high velocities, even when using high-speed cameras. One example is the measurement of self-speed from images of the road surface taken with a camera from a driving car. In view of this purpose we explore the potential of the method and its limitations.

### Keywords

Optical flow, high speed, noise, low resolution

## 1. INTRODUCTION

Optical flow is, both in biology and in technology, an important phenomenon. It is the motion of structures in a two-dimensional projection from a three-dimensional scene. In biology, optical flow is considered crucial for animal and human vision, the detection of moving objects and the experience of self-speed. In computer vision, optical flow is the basic observable to quantify speed of objects relative to other objects, or relative to the camera taking the scene. The latter process is referred to as self-speed estimation or measurement.

In this paper we research the feasibility of self-speed measurement of a car driving on a highway based on optical flow, under poor visibility conditions such as low-contrast texture. In section 2 we describe the motivation of the investigation and the requirements of the method. Also the scope of this paper will be precisely defined. In section 3 we present the technical approach and in section 4 we define the experimental conditions. In section 5 we present the experiments and results. Conclusions about the feasibility of the method and outlook are described in Section 6.



## 2. MOTIVATION AND REQUIREMENTS

The research aims at a method to extract and record self-speed of a car and in addition visual scene information, using an instrument which can be easily mounted in a car. This paper deals with, and its scope is limited to, the feasibility of such instrument. The background of the research is the study on car drivers' behavior during highway traffic congestions and on new approaches to influence the driving style and to reduce traffic jams. Both the emergence of congestions and their dissolution have got due attention in research over decades. Based on traffic flow models, simulation systems have been developed to study all kind of phenomena of traffic on micro and macro scales, under various conditions in various scenarios. The behavior of individual car drivers is central to most traffic flow models, in particular when they categorize as car-following models. These models assume that a car driver controls his/her car mainly as a function of other cars, in front (if any). The acceleration (positive or negative) of the car is modeled as a function of own speed and the speed of the car ahead and its distance.

Treiber (2003) proposed that car-following characteristics not only differ among drivers, but may also vary for one driver over time. He has modeled memory effects in the response behavior of drivers to the traffic situation, *e.g.* by presuming that after being stuck in a jam, drivers tend to increase the time gap to the preceding vehicle. When incorporating this memory effect into the IDM (Intelligent driver model), traffic flow simulations get increasingly consistent with the measured data (Treiber 2003).

If cars tend to accelerate slowly after leaving a traffic jam, it may cause a significant decrease of road capacity and cause congestions to be over-persistent. Numerical simulations have shown that the life time and length of jams as well as delays of individual cars increase significantly when the acceleration parameters of the IDM are reduced [Vergeest 2012]. Although there are anecdotic indications that people tend to leave traffic jams too slowly, we have not found any objective reporting about this. It is the main goal of our research to obtain statistical information about drivers' behavior just after having been stuck in a jam. Without this information being available, we can only speculate about ways to influence and improve the driving style and thus to avoid OPCs (over-persistent congestions).

We should point out here that the ACC (adaptive cruise control) and similar systems could reduce the problem of OPCs. However, although the penetration rate of ACC is increasing, it is not yet evident that they operate efficiently and safely during congestions and other low-speed situations (Xiong 2012). Therefore we will focus on fully human-controlled cars.

One could reflect about obtaining statistics of human driving behavior by using a car simulator. Suppose that the car simulator were based on the IDM (or similar) model. Then the virtual traffic provided by the simulator accelerates according to the IDM. The subject's car is operated by a test person, allowing the actual acceleration characteristics (and other parameters) of the test person to be recorded. The scenario provided by the simulator may contain congestion conditions. In this way, using the data from many test persons, the acceleration profile (as function of distance and speed) could be statistically obtained. There is, however, one basic assumption undermining this approach. The virtual, surrounding traffic generated by the simulator is based on the IDM and not on the actual acceleration profile, which is actually the unknown we are after. The difference between the model's profile and the real-life profile could bias the profile exhibited by the test persons driving the simulator.

In our aim to obtain the real-life acceleration behavior of drivers we focus on three main parameters: 1) time, 2) the speed of the own vehicle and 3) the distance to the car ahead. From the latter the speed of the car ahead (relative to the own car) can be derived. Although useful, the location of the own vehicle as a function of time is not needed to detect the occurrence of OCPs. We limit our study to cars in a single lane. Our interest is in situations were cars leave a congestion, which represent, however, only a small fraction of typical journeys.

One way to collect sufficient statistics is to record the three aforementioned parameters of cars participating in traffic over long periods of time. The situations of interest should then be filtered out in subsequent data analysis.

Once somebody volunteers to participate in the research and, it should be made very easy to adapt his/her own car. We formulate the requirements of the data taking system:

1. It should be portable and easily and quickly installable in any common passenger car. The car driver him/herself should be able to install and take out the system in less than 1 minute.
2. It should require no or very little effort or attention to operate the system. A single on/off switch should suffice. Automatic switch on/off is also an option. It should not distract the attention of the driver during driving.
3. The instrument(s) should not be expensive or otherwise attract the attention from people passing by.
4. The recording capacity should be sufficient for about 50 to 100 hours of driving time.

We list no requirements about computer processing, assuming that the data needs not be analyzed real-time. The extraction of the three parameters from the raw data will be done offline as will be the analysis.

Let us define the scope and purpose of this paper more precisely. The main purpose of this paper is to find out whether it is feasible to collect empirical data subject to the 5 requirements listed above. (The conclusion of the paper is that it is not, to our best knowledge). In Section 3 we reason that a possible approach could be based on a simply mountable camera, inside the car, viewing into forward direction. To actually detect optical flow from image sequences can be done with a multitude of methods, as *e.g.* reviewed by [Barron 1994]. Although the performances of the various analysis techniques differ, their effectiveness in view of our application is outside the scope of our paper. The major factor that limits the feasibility seems to be the image quality, rather than analysis performance. Therefore, our main focus is on requirements of image resolution at high self-speed.

# 3. TECHNICAL APPROACH

The most direct way of recording speed is to simply readout the car's own speedometer. In principle, the speed (as many other data) can be obtained in digital form via a cable connector provided by the car manufacturer. However, the connection and the interface are not standardized, which makes it unpractical, considering requirement 1.

The recording of own speed as a function of time could also be done using a GPS logging device. However, deriving your speed from GPS coordinates and time is sensitive to the accuracy of the GPS coordinates, which is know to vary depending on the quality of and the number of the satellite signals received. Some GPS devices can measure speed directly from the Doppler shift, which is more accurate, but also dependent on the satellite reception quality. In practice, measuring time and own speed using GPS is a good option, and might meet all 4 requirements.

However, we also need a practical way to measure the distance to the car ahead as a function of time. If one would have access to the GPS data from own car and from car ahead, the problem would be solved. In a controlled field experiment, it proved possible to collect data from 50 cars on a real highway, which was reserved for the time of the experiment [Schakel 2010]. Each of the cars was equipped with a cooperative ACC, where the ACCs could communicate among each other. Theoretically, when all cars on all motorways would measure and communicate position, then the distance between any two cars could be obtained, see *e.g.* [Herrara 2010]. In practice this is not yet feasible.

We are therefore looking for a way to measure the distance to the vehicle ahead from our own car, as a function of time. We consider two options. The first is radar-based. This concept is central to ACC systems and works reliably. It conflicts, however, requirement 1 and, since it would measure distance only, the recording of speed would involve one additional device or instrument. The other option is vision. As demonstrated in [Nieto 2010], the detection of own speed and of the distance to cars ahead can be retrieved from footage from a single onboard video camera. The measurement of speed is reliant on the detection of optical flow, or the displacement of image features between two subsequent frames from the video camera. Relatively clear image features are white lane markings on a dark surface road. However, at high velocity, the displacement of lane markings may become as large as the distance between the lane markings themselves, which is complicating the computation of speed [Vergeest 2012a]. Although lane markings are typically present on highway road surfaces, speed measurement should not depend on their occurrence.

In general, temporal aliasing effects deteriorate the performance of optical flow methods [Marmarella 2012]. However, also poor light conditions, camera noise and motion blur in the individual images pose a problem to the derivation of speed from the images.

However, suppose that we could solve these problems, then we have an instrument that would fulfill all 4 requirements. The instrument is an onboard camera providing images of the scene ahead, at a certain frame rate. Even when the images do not contain a time stamp, knowing the time difference between subsequent images is sufficient for our purpose. From the optical flow we can derive the forward speed of the own car, not necessarily to be determined from lane marks for in case these are absent or too far apart, but in some other way, still to be found. Furthermore, from the same images we can detect the distance of the car ahead, based on Nieto's method or as in [Vergeest 2012a].

We remark that optical flow detection should not necessarily depend on the availability of image frames. If the photo sensors of the camera could be read individually and instantly, the motion of image features might be even better identified.

In conclusion, as a technical approach we consider an onboard camera. However, as the most dominant problem, we need to address self-speed estimation from optical flow at high velocity and poor visibility conditions in absence of clear features such as lane markings.

# 4. EXPERIMENTAL CONDITIONS

In this section we study the feasibility of estimating self-speed from optical flow. The main principle is to determine the displacement of objects in two subsequent images. We have discarded the rotational component, which can be expected to be small compared to the translational component, at high velocity. Also possible bends or slopes of the road surface are not taken into account, assuming that their effects will be small.

As mentioned, when the scene contains clearly detectable features which can be well correlated, the displacement can be reliably determined, and the speed of the object relative to the camera be estimated [Souhila 2007]. In our application we focus on the road surface. Assuming that (at least locally) the road surface is consistent with a plane parallel to the driving direction of the car, points in the surface can be one-to-one mapped from 3D to points in the image plane of a camera fixed onto the car. Although the optical flow of objects such as trees or buildings may be relatively easily determined in the image plane, there would be no simple one-to-one mapping to the 3D scene.

When the road surface does not contain obvious features, such as lane markings, the visual structure or texture of the road surface might be used to detect the amount of shift as a function of time as a measure of optical flow. When two images of the road surface are available, taken from the same camera at a small time difference, the matching criterion can be defined as a difference function of the two images, which should be minimized [Mammarella 2012]. At low speed, when the amount of shift is small, it is relatively easy to find matching regions in the two images. At high speed, there are two problems of finding matching image regions. The first is that the expected shift will be large, and therefore a larger range of potential shift needs to be considered. The probability that nearly similar structures pop up will increase, and so will the risk of false matches. Second, due to the high speed the image quality will degrade, either due to motion blur, to shorter shutter times (and hence an increased camera noise level), or both. When reducing the first problem (increasing the frame rate) the second problem gets worse.

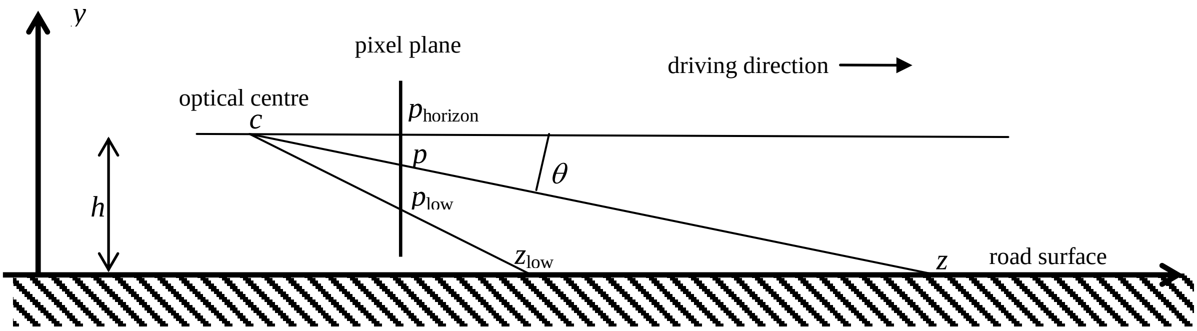


**Figure 1. Perspective parameters of the camera setup. The optical center *c* is at height *h* above the road surface. The driving direction is into the *z*-direction.**

Let's consider the simplified camera setup in Figure 1. The camera is mounted inside the car near the front window, pointing into the driving direction, taking images as in Figure 2. In the pixel plane of the camera, which is assumed to be vertical, $p_{horizon}$ is the index (counting from bottom up) of the pixel line representing the horizon. $p_{low}$ is the lowest pixel line showing road surface. That particular line in the road surface has *z*-coordinate $z_{\text{low}}$ as measured in the coordinate frame with origin *c*. The height of the optical center *c* has *y*-coordinate equal to *h*. A point in the road surface at distance *z* from the camera in forward direction will be mapped to a pixel on scan line *p*, such that

$$z = z_{\text{low}} \frac{p_{\text{horizon}} - p_{\text{low}}}{p_{\text{horizon}} - p}. \qquad (1)$$

We define $\Delta p$ as the amount of vertical shift observed for a point on the surface between two successive camera pictures. $\Delta p$ depends on the speed *v* of the car, the resolution and the frame rate of the camera, and on the location of the point in the perspective image. For points projected near the bottom of the picture, the optical flow measured in pixel shift is relatively large. For large $\Delta p$ the image matching process will be computationally more involved and the risk of error will increase. Therefore, $\Delta p$ will be an important parameter for the trade-off between image quality and maximum speed.

**Figure 2. Picture taken with an onboard camera. The region of analysis is indicated by the white box below the center of the image.**

In the exploration of the feasibility of image matching, we will set some further limitations to our scope. Out of the various types of difference functions we chose to use the sum of absolute differences (SAD) among pixel brightness of the image regions. Another assumption is that the image regions are located near the plane *x*=0 of the camera, as to simplify the compensation for perspective distortion.

The focus of initial experimentation will be the ability to determine $\Delta p$ from two given images, where we assume that 1) the images have been taken by the same camera, 2) a predefined region in the image is of interest, 3) the image in the region of interest is created according to the setup as in Figure 1.

## 5. EXPERIMENTAL RESULTS

We collected images with an HD-HERO2 video camera from GoPro [Gopro 2013]. This camera can easily be installed as a dashboard camera. We setup the camera at frame rate $u$ =120 fps. The field of view was 170° at a resolution of 848×480 pixels. The optic flow of the road surface was determined from the pixel data in a small region of the images near $z$ = 4.2m, see Figure 2. In this particular case the region is of size 180×32 pixels. The SAD is determined by moving a subwindow (smaller than the region) from one image over the region of the second image in steps of one pixel. In the current algorithm we ignore the (small) perspective distortion present in the regions. Suppose that we choose the subwindow to have size 160×12 pixels. Then the maximum number of steps in the $Y$-direction is 32-12=20, that is 9 pixels in each direction, which is the maximum $\Delta p$ that would be detectable.

For the camera setup we have $z_{low}$ = 4.0m, $p_{low}$ = 150 and $p_{hor}$ = 320. The mapping factor $f$ is the distance on road surface corresponding to an increment of one scan line,

$$f = \frac{\partial z}{\partial p} = \frac{z^2}{z_{low}(p_{hor} - p_{low})}, \qquad (2)$$

as can be derived from equation (1). At the bottom of the image region, which is near $z = z_{low}$, a shift of one pixel into the vertical direction corresponds with approximately $f$ = 4.0/(320-150) = 0.024m distance on the road surface for our specific camera setting. Since $\Delta p = v / uf$, we have $\Delta p$ = 3.5 pixels at $v$ = 10m/s, or $\Delta p$ = 11.6 pixels at 120km/h.

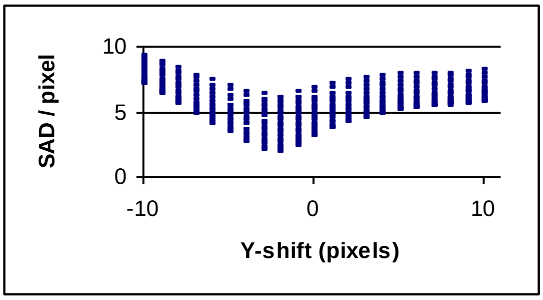


**Figure 3. SAD of two consecutive images as a function of *Y*-shift. For each *Y*-shift 21 points are shown for *X*-shift = -10, -9, ...., 9, 10 pixels.**

Figure 3 shows, as an example for one pair of images, the SAD (divided by the number of pixels in the sub window) as a function of Y-shift. In this case we observe a minimal SAD at Y-shift or $\Delta p \approx$ -2, corresponding to $v \approx$ 6m/s or 21km/h. The lowest point in the plot of Figure 3 corresponds to an X-shift of 0 (not visible in the plot). The speed derived from the plot is $v = u\ f\ \Delta p$, that is it is therefore discrete in steps of $u\ f$ = 2.9m/s or 10.4km/h. One could attempt to fit a curve to the minimum SAD as a function of Y-shift and thus estimate $v$; we have not done that.

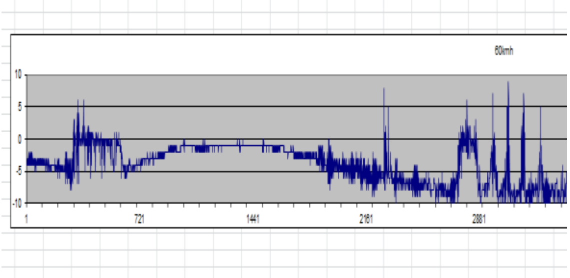


**Figure 4. $\Delta p$ as a function of time over a 30s time interval.**

The footage from the HERO2 allows the measurement of $v$ at its frame rate, provided that we can determine $\Delta p$. An impression of the reliability to determine $\Delta p$ is given in Figure 4. The course of $\Delta p$ can be recognized, but there are some “noisy” parts, which correspond to locations on the road where the surface lacks visual contrast. Figure 5 provides more detail about this effect.

In Figure 5 we show the measured speed as a function of time over a 1 second time interval. The hosting car was driving at a constant speed of approximately 30km/h. In the plot the speed is presented in units of $\Delta p$. The SAD and the mean AD (the absolute differences averaged over all computed shifts) are included in the figure as well, where AD ≥ SAD. For 10.4 < $t$ < 11.1s $\Delta p$ takes the value -2, which is quite consistent with the actual speed. Outside the interval, $\Delta p$ seems scattered. Where both AD and SAD are small, $\Delta p$ cannot be determined reliably; the image pairs do not exhibit enough contrast. Figure 6 depicts the scene where the change of road surface texture from rough to smooth occurs, near $t$ = 11.1s.

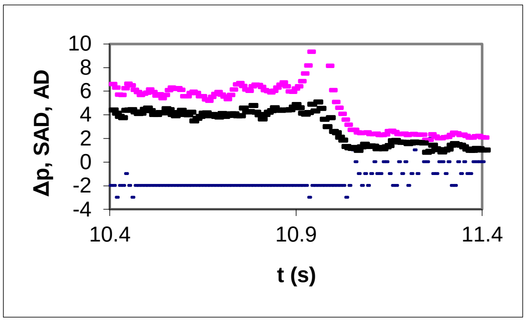


**Figure 5. $\Delta p$, SAD and mean AD as a function of time, over a one second time interval.**

**Figure 6. Road surface texture changes from high to low contrast at *t* = 11.1s.**

When a smooth road surface exhibits little contrast, the SAD as function of *Y*-shift (Figure 7) is very different from the profile obtained in Figure 3. In Figure 7 we notice that both the SAD values themselves as their variation due to shift are much smaller.

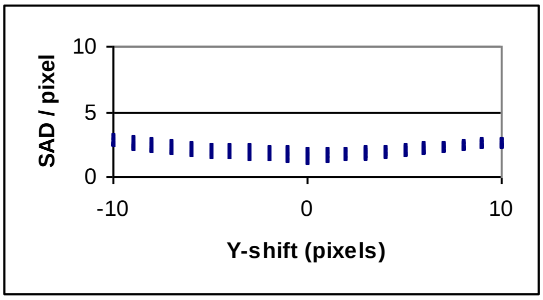


**Figure 7. SAD for an image pair of low contrast road surface. For convenience of comparison, the plot scales are the same as for Figure 3.**

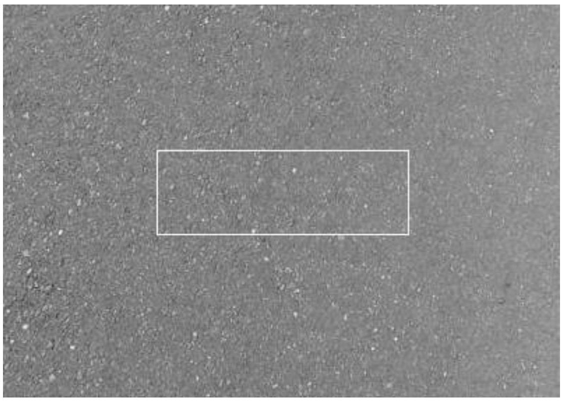

**Figure 8. Part of the low-contrast road surface (also visible in Figure 6), taken at higher resolution. The dimensions of the view are approximately 40×45cm. The white box represents the portion of the image which is used for view matching.**

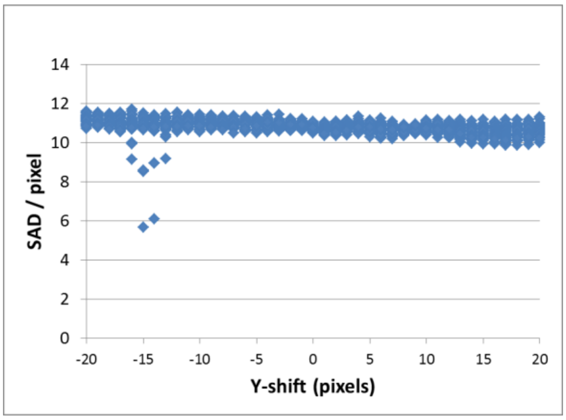


**Figure 9. SAD for an image pair of low contrast surface, where images are taken at high resolution.**

As mentioned, there are several parameters that may influence the SAD profile, such as the frame rate of the camera, its resolution and shutter time, but also parameters of the software including the choice and size of image region and subwindow.

Concerning the image contrast itself, for now we consider two criteria. First, the degree of contrast that is required to achieve matching. Second, the degradation of contrast due to the speed of optical flow.

With a photo camera we took a detailed still picture of the road surface in the low-contrast region, see Figure 8. Another similar picture was taken after the camera was manually repositioned a few centimeter further in the positive *z*-direction. Whereas the particular part of the surface road did not show contrast in Figure 6, it does in Figure 8.

From two detailed images we could reliably find a match near $\Delta p$ = -15, see Figure 9. The *Y*-shift in Figure 9 maps to a speed relative to the road surface as by equation (2), where *f* differs from the mapping factor we applied so far since Figure 9 has been obtained with a different camera setup.

We thus found that the speed of the car could be determined from detailed pictures as in Figure 8, even when the image contrast is low. However, the picture pair from which Figure 9 is derived would not be reproducable with the video camera we applied earlier, for two reasons. First, due to the high resolution, the optic flow measured in pixels/s would be very high (in the order of 3000 pixels/s) and hence a very short shutter time would be required to obtain a non-blurred image. Second, the frame rate should be an order of magnitude larger, to keep the *Y*-shift between the two images between 10 and 20 pixels. This latter requirement is of course depending on the maximum speed we intend to measure.

## 6. CONCLUSIONS

If a road surface exhibits sufficient visual contrast, the optical flow can be captured with simple equipment at low cost. The collection of large statistics data over long traveling times about the driving speed of a car (and its distance to cars in front) would then be feasible. We have demonstrated that the footage from a simple video camera is sufficient to measure the car's speed without being dependent on artificial features such as white road markings.

When the surface road contrast gets low and/or the car speed gets high, the method becomes unreliable. However, even low-contrast asphalt exhibits texture from which optical flow can be detected, provided that high-quality images at small time intervals were available. It would not be necessary to continuously store images at a high frame rate. For our experimental research it is sufficient to obtain (for example) only one image pair per second, where the time difference between the images is small (perhaps in the order of 0.1ms or less). We have not yet found a device which could perform like this.

A possible method could be to apply two photo cameras. The cameras should be mounted closely adjacent, aiming at the same point on the road, probably highly zoomed. Shutter times should be short. Then the cameras should be triggered to take one picture every second, where one camera is triggered 0.1ms later than the other one. If pictures as in Figure 7 are obtained this way, the optical flow and hence the speed of the car relative to the ground can be recorded as a function of time.

There is another advantage of applying higher resolution, and thus small $f$, considering the definition of $f$ in equation (2). The speed is derived from the $Y$-shift, which is essentially a discrete value, although the appearance of plots as in Figure 3 suggest that a curve can be fitted against the data, from which a minimum could be derived. If the discrete minimum is used, the speed $v = u\ f\ \Delta p$ is measured in steps of $uf$. Since the frame rate $u$ should be high, $f$ should be as small as possible in order to determine $v$ accurately.

Another factor is the luminous intensity of the scene. If high intensity spotlights are applied, a small $f$ can be reached. It would involve the installing of extra high-power lights on a car, which violates requirement 1.

As mentioned, the detection and quantification of optic flow does not necessarily require image frames. If individual pixels of the optical sensor could be read-out at high speed, for example 10KHz, then even small and low-contrast features might be traceable. Their speed in sensor space would be a measure of the car's speed. A similar principle is applied in the optical mouse.

Another option could be the projection of a laser beam onto the road surface during a predefined time period $x$. A picture of the road surface taken with a relatively long lens opening time (much longer than $x$ seconds) will show a line on the road surface caused by the laser beam. The length of the line, which can be reconstructed using equation (2) is proportional to $v$. The advantage of this method is that low-quality cameras could be applied. However, the total setup of equipment gets more complicated due to the inclusion of the laser device.

At present, the most feasible method (fulfilling the 4 requirements) seems to be based on optical flow detection using a single camera or a synchronized pair of cameras. It is still an open question whether images with quality comparable to the one in Figure 8, can be obtained from a car driving at high speed on a road with low-contrast surface texture. Weather and light conditions play an important role as well.

## 7. REFERENCES

[Barron 1009] J.L. Barron, D.J. Fleet, S.S. Beauchemin, Performance of Optical Flow Techniques. Int. Journal of Computer Vision, 12:1, pp 43-77.

[Gopro 2013] www.gorpo.com.

[Herrara 2010] D.B. Work, R. Herring, X. Ban, Q, Jacobson, A.M. Bayen, Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. Transportation Research Part C (2010), pp 568-583.

[Mammarella 2012] M. Mammarella, G. Campa, M.L. Fravolini, M.R. Napolitano, Comparing optical flow algorithms using 6-DOF motion of real-world rigid objects. IEEE Trans on Systems, Man and Cybernetics – Part C, Vol. 42, No 6, pp 1752-1762.

[Nieto 2010] M. Nieto, J. Arróspide Laborda, L. Salgado (2010), "Road environment modeling using robust perspective analysis and recursive Bayesian segmentation". Machine Vision and Applications, 7 August 2010.

[Schakel 2010] W.J. Schakel, B. van Arem, Effects of cooperative adaptive cruise control on traffic flow stability. Proc. 13th Int. IEEE Annual Conference on Intelligent Transportation Systems, pp 759-764.

[Souhila 2007] K. Souhila and A. Karim, "Optical Flow based robot obstacle avoidance," Int. J.

Adv. Robot. Syst., vol. 4, no. 1, pp. 13–16, 2007.

[Treiber 2003] M. Treiber and D. Helbing, Memory effects in microscopic traffic models and wide scattering in flow-density data. Physical Review Letters E 68, 046119 (2003), pp 1-8.

[Vergeest 2012] J.S.M. Vergeest and B. van Arem, The effect of vehicle acceleration near traffic congestion fronts. In 2012 IEEE Intelligent Verhicles Symposium IV'12, pp. 45-50.

[Vergeest 2012a] J.S.M. Vergeest, Self-speed and headway measurement in highway traffic from onboard video footage. In V. Skala *et al.* (Eds.) Proc. of the WSCG 2012 Conference.

[Xiong 2012] H. Xiong and L. Ng Boyle, J. Moeckli, B.R. Dow and T.L. Brown, Use patterns among early adopters of adaptive cruise control. Human Factors, Vol. 54, No. 5, October 2012, pp.722-733.