ROBUST VEHICLE TRACKING IN VIDEO IMAGES BEING TAKEN FROM A HELICOPTER

Fatemeh Karimi Nejadasl, Ben G.H. Gorte, and Serge P. Hoogendoorn

Institute of Earth Observation and Space System, Delft University of Technology, Kluyverweg 1, 2629 HS, Delft, The Netherlands
f.KarimiNejadasl, b.g.h.gorte@tudelft.nl
Transport and planning section, Delft University of Technology, Stevinweg 1, 2628 CN, Delft, The Netherlands
S.P.Hoogendoorn@tudelft.nl

Commission VII

KEY WORDS: Optical Flow, Tracking, Feature detection, Matching, Region based, Feature based

ABSTRACT:

Measuring positions, velocities and accelerations/decelerations of individual vehicles in congested traffic with standard traffic monitoring equipment, such as inductive loops, are not feasible. The behavior of drivers in the different traffic situations, as re-quired for microscopic traffic flow models, is still not sufficiently known. Remote sensing and computer vision technology are recently being used to solve this problem. In this study we use video images taken from a helicopter above a fixed point of the highway.
We address the problem of tracking the movement of previously detected vehicles through a stabilized video sequence. We combine two approaches, optical flow and matching based tracking, improve them by adding constraints and using scale space.
Feature elements, i.e. the corners, lines, regions and outlines of each car, are extracted first. Then, optical-flow is used to find for each pixel in the interior of a car the corresponding pixel in the next image, by inserting the brightness model. Normalized cross correlation matching is used at the corners of the car. Different pixels are used for solving the aperture problem of optical flow and for the template matching area: neighboring pixel and feature pixels. The image boundary, road line boundaries, maximum speed of the car, and positions of surrounding cars are used as constraints. Ideally, the result of each pixel of a car should give the same displacement because cars are rigid objects.

1. INTRODUCTION

Traffic congestion is an important problem in modern society. A lot of money and time is wasted in traffic jams. Car crashes and accidents are more frequent during busy traffic conditions. Several efforts are made to tackle this problem: better facilities and regulations should improve the situation on existing roads while the number of the roads is extended as well.

Traffic congestion is highly dependent on the behavior of individual drivers. For example, reaction times and lane-changing techniques vary from driver to driver. Therefore it is useful to model the behavior of individual drivers, as well as the interaction between drivers, before new decisions and regulations for traffic congestion control are initiated. Current traffic theories are not yet able to correctly model the behavior of drivers during congested or nearly congested traffic flow, taking individual driver's behavior into account. For this so-called microscopic traffic models are needed. Vast amounts of data are required to set up those models and determine their parameters.

Traffic parameter extraction with airborne video data is recently getting popular. Automatic extraction of traffic parameters is a computer vision task. For traffic parameter extraction, information about each vehicle is needed during the period of time the vehicle is present in the scene. A possible solution is to detect a vehicle in a video frame when it enters the scene and then track it in successive frames.

The video is recorded by a camera mounted on a helicopter. Since we want to model the behavior of as many vehicles (drivers) as possible, we attempt to cover a large highway section, leading to the lowest spatial resolution that accuracy requirements allow. Typically we use a spatial resolution (pixel size) between 25 and 50 cm.

Helicopter movement invokes camera motion in addition to object (i.e. vehicle) motion. We have removed camera motion with the method describes in (Hoogendoorn *et al.* 2003) and (Hoogendoorn *et al.* 2003). Unwanted areas outside the road boundary are eliminated by (Gorte *et al.* 2005).

In earlier work, vehicles were detected by a difference method (Hoogendoorn *et al.* 2003), which requires involvement of an operator when automatic detection fails. This is often the case with cars having low contrast against the background (dark cars on a dark road surface). We used cross correlation matching for tracking. This works well in the case of distinct features with homogeneous movements. In this case it is less sensitive to the illumination change. However it is too sensitive to similarities in texture or brightness.

To improve the performance of tracking, we investigate the use of *optical flow* methods in this paper. Improvement with respect to least square matching (Atkinson 1996) is expected because of the additional time element in the optical flow equation.

Optical flow method is sensitive to small (even sub-pixel) movements. This sensitivity may be helpful for tracking cars that are similar to the background.

The paper is organized as follows. In section 2. we present related work. Section 3. discusses zero cross correlation matching method, in section 4. gradient based optical flow method by assumption of constraint and linear model of brightness is discussed. Feature selection and constraints are described in the result redundancy exploitation section. We give results in section 6. and conclusions in section 7..

2. RELATED WORK

Automatic object tracking receives attention in computer vision for a very diverse range of applications.

Matching methods are largely used in video tracking. As mentioned earlier, they are quit good in distinctive objects. However

they have a problem with repeated patterns, areas with similar intensities and very large displacements, all of which occur in car tracking from helicopter sequences. Optical flow is the alternative way in tracking. However, it is a challenging task in our sequence with different motion sources.

(Haussecker and Fleet 2001) described physical model for different brightness variation. They improved the optical flow equation in the different brightness condition, linear and nonlinear. While using complicated model is more suited for a big object because of increasing the number of parameters.

There are two different methods for tracking: feature based and region based. The feature based method have some advantages in case of occlusions and changing intensities. Even if some part of an object is occluded by another object, remaining visible points may still be detected. (Smith and Brady 1995), (Smith 1998) and (Coifman *et al.* 1998) used feature based methods in vehicle tracking. Region based methods preserve the shape of objects and the chance of wrong correspondences is decreasing. However there is a strong possibility to loose the tracking in occludes areas, which may occur in dense traffic. (Cohen and Medioni 1999) proposed a graph representation of moving objects in the tracking of regions.

(Partsinevelos *et al.* 2005) presented a method called ACENT (Attributed-aided classification of entangled trajectories) for solving the ambiguity in tracking specially in the entangling of different trajectories together.

The approach presented in this paper modifies highway vehicle tracking with combination of different results. The results of the different methods of matching and optical flow considering constant or linear variation of brightness are cooperate the final result. Different features (pixels inside the car and corner using neighbors the car boundary and region) are used in different method. Therefore redundant results are provided. Using different constraint such as maximum speed, image and road boundary, and neighboring cars suppress errors. Initial value improves the results.

3. ZERO NORMALIZED CROSS CORRELATION

This method is based on the searching of a template window (the neighboring window around of a specific pixel in the first image) in the searching area in the second window. For each pixel in the searching area, neighboring pixels make a new window with the same size of template window. The below-equation is used to calculate correlation between two template area in two successive images:

$$\rho = \frac{\sum_w [I_{i1}(x,y)-\overline{I_1}][I_{i2}(x-u,y-v)-\overline{I_2}]}{\sqrt{\sum_w [I_{i1}(x,y)-\overline{I_1}]^2 \sum_w [I_{i2}(x-u,y-v)-\overline{I_2}]^2}} \quad (1)$$

which w shows the neighboring-pixel positions in the first image, u and v are displacement of specific pixel. For the simplicity of zero normalized cross correlation (ZNCC), $I(x_1(i), y_1(i))$, which shows the brightness of the first image for the point i, is replaced with $I_{i1}(x, y)$. In the similar way, $I_{i2}(x, y)$ shows the brightness of point i in the second image.

The maximum ZNCC indicates the best match of the specific pixel.

ZNCC is a suitable method for tracking of distinctive features. There is an ambiguity in the area with similar brightness or similar texture. Therefore only corner points are tracked with this method.

4. OPTICAL FLOW METHOD

A proposed method by (Lucas and Kanade) is used to calculated the displacement of each detected point in the different frames. It is assumed that the brightness of a pixel belonging to a moving feature or object is remaining fixed in consecutive frames. This assumption is mathematically translated to the below form:

$$I(x_1, y_1, t_1) = I(x_2, y_2, t_2) \quad (2)$$

In the above equation, I, x_i, y_i, and t_i denote brightness, spatial coordinate and time in the first and second image frame. The Taylor series expands the above equation to the spatiotemporal elements but only the first order is used. It is assumed that there is only translational movement. Therefore we rewrite the equation into the form of gradient elements as:

$$\begin{gathered} I(x_1, y_1, t_1) = I(x_1 + u, y_1 + v, t_1 + \mathrm{d}t) \\ I_x u + I_y u = -I_t \end{gathered} \quad (3)$$

where u and v are displacement values in x and y directions respectively. We also call them the optical flow parameters. I_x, I_y, and I_t refer to the gradients in spaces and time. Equation 3 is the well-known optical flow constraint equation (OFCE) or brightness change constraint equation (BCCE).

The OFCE is simplified to the non-linear observation equation (Teunissen 2003), (Jahne 2001):

$$\underline{y} = A(x) + \underline{e} \quad (4)$$

, or

$$E(\underline{y}) = A(x) \quad (5)$$

in which $A = \begin{bmatrix} I_x & I_y \end{bmatrix}$, $\underline{y} = \begin{bmatrix} I_t \end{bmatrix}$, and $x = \begin{bmatrix} \hat{u} & \hat{v} \end{bmatrix}^T$ are respectively coefficient, observation, and parameter matrices.

The only parameters of the OFCE are translations, i.e., optical flow parameters. At least two pixels are required for solving the equation. All pixels used for solving the equation for a specific pixel are assumed to have the same displacement vector (the same optical flow parameters) as that specific pixel.

In the OFCE, space and time gradients (I_x, I_y, and I_t) play the important role of constructing the coefficient and observation matrix.

The gradients in are calculated from two consecutive images in order to include both time and space in Equation 3. The below-convolution matrices are respectively used to calculate the space, x and y, and time gradients:

$$C_x = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}, C_y = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}, C_t = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Figure 1 shows the displacement results using optical flow assuming the constant brightness model. The amount and direction of each displacement is depicted by the magnitude and direction of the arrow.

Initial values are required to correctly calculate the optical flow parameters. With the good estimation of the initial value, the chance to find an incorrect correspondence is decreasing.

The area which is used to calculate the gradients and then the optical flow parameters is updated by the initial values.

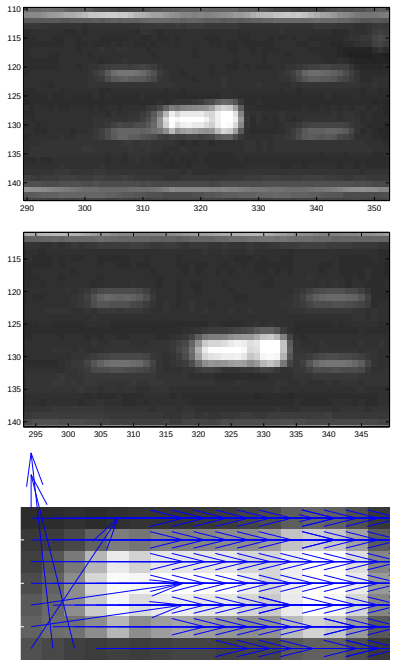

Figure 1. Displacement of each inside-car pixel in between two consecutive frames: the first and the second row show two consecutive images; the displacement result for each pixel is presented by an arrow

Firstly, the optical flow parameters are calculated in the most coarse resolution. Then they are again calculated in a finer resolution using the scaled results of the previous level. This process is continued until the original scale is reached.

When the displacements between pixels in the different successive frames are large, the chance of getting wrong results is increasing. Using the results from a coarser scale improves the results in the following step and reduces the chance of errors.

The following algorithm describes the tracking process which is used in our implementation.

```
for frame = 1 to Nframes do
  for scale = coarse to fine do
    repeat
      for feature type in {points, corner,
        boundary, region} do
    until the observation error is less than a
     threshold
  end
end
```

The result of each pixel is independent to the results of the other pixels. With this approach, a wrong result can not infect the other results.

4.1 LINEAR BRIGHTNESS MODEL

Using a linear model is suggested in the case of changing brightness. Consequently the linear brightness model is substituted for the OFCE. This is translated mathematically as follows:

$$I(x_1, y_1, t_1) \;=\; aI(x_2, y_2, t_2) + b \qquad (6)$$

$$I(x_1, y_1, t_1) = \; a(I(x_1, y_1, t_1) + aI_x u + aI_y v + ag_t + b) \quad (7)$$

if both sides of the equation are divided by a one obtains:

$$I_x u + I_y v + I_t + \frac{a-1}{a} I(x_1, y_1, t_1) + \frac{b}{a} \;=\; -\, g_t \qquad (8)$$

The parameters $\frac{a-1}{a}$ and $\frac{b}{a}$ are changed to k_1 and k_2 respectively.

$$I_x u + I_y v + I_t + k_1 I(x_1, y_1, t_1) + k_2 \;=\; -\, g_t \qquad (9)$$

Non-linear least square approach (Teunissen 2003) similar to the OFCE is used to solve the above equation by determining 4 parameter values u, v, a, and b compared to the above. Just the coefficient matrix, $A = \begin{bmatrix} I_x & I_y & I & 1 \end{bmatrix}$, and parameter matrix, $x = \begin{bmatrix} \hat{u} & \hat{v} & k_1 & k_2 \end{bmatrix}^T$, are changed.

5. RESULT REDUNDANCY EXPLOITATION

None of the above-described methods can give the correct results independently in the case of changing brightness and shape. Therefore in this paper we try to provide the correct results as much as possible for each vehicle.

According to the Equation 9 and OFCE, only a few pixels are required to solve these equation. These pixels should have the same displacement. In the conventional method the neighboring pixels in the specific window area are used to solve the equation OFCE. Here we have used another pixel as well. In the ideal case the results should be the same using different pixels. But because of unavoidable errors occurring in brightness and shape variations the results are in general not the same. Therefore we should decide whether a solution is correct or incorrect. Finally we get a correct value for the displacement using the redundancy in the results.

The objective of this section is to describe how to provide the different results obtained using different methods, area based matching and optical flow, as well as using different pixels.

5.1 FEATURE SELECTION

Extraction of car features is required to prepare the redundant exploitation of results. These features are car pixel, region, boundary, and corner.

Displacement of each pixel inside the car is calculated using optical flow. The OFCE is provided by the optical flow elements. The 3×3 neighboring pixels are participating in the OFCE to calculate the displacement of the central pixel.

The car region is extracted using a reference image (Hoogendoorn *et al.* 2003). The first frame is reduced from the reference image being made median of whole frames. The car region is obtained by a threshold and using morphological operations. However the whole vehicle can not be detected by this method. It is highly dependent on the reference image and the selected threshold. Instability of the helicopter and variable weather conditions change the brightness of the road surface and the road lines, even in successive frames. Due to the brightness changing of fixed objects and similarity between dark car and the road surface brightness, a small threshold inserts a lot of errors and while a big threshold causes the dark cars to get lost. In this paper the dark cars are detected manually. The fully automatic detection of vehicles is suggested for the future.

The car boundary is extracted from the region by a morphological operation. The extracted car-regions are eroded by structural

element with all array one and then is reduced from the region image. The result is produced the car boundaries.

Harris method (Harris and Stephens 1988) is used to detect points automatically inside the road boundary. Using set theory removes the points which do not belong to the car region. The only car points are accepted as the final results. Manual detection is extracted the rest of the car-corner points which are not extracted by automatic method.

As it described above, the different pixels are used to solve the OFCE. The tracking by the region and boundary pixels is working in the similar way. They preserve the shape as well as the brightness characteristics. Therefore the results are more reliable.

To avoid of complexity of gradient calculation, both images are convolved to C_x, C_y, and C_t. Then in each image according to the position of selected pixels the convoluted results are extracted and combined.

In iteration only the position of pixels in the second image and thus their gradient are updated.

In the other features, corner and inside pixels, because of using neighboring pixels, the fast and easiest way of calculating gradients is convolution of C_x, C_y, and C_t in only this region and then combination of them.

The region are extended one pixel in all sides for correct calculation of gradient. After the whole gradient calculation, one pixel from all sides are removed. In the same way as boundary and region, only the pixels of the second image are updated.

5.2 CONSTRAINT

The wrong results especially because of the similar brightness or texture are discarded using boundary constraints. The road and image boundary, maximum speed of car and speed of neighboring cars are used as the boundary constraints.

The quality of data also is determined before calculation of results. $\left|A^T A\right|$ should not be zero otherwise the equation OFCE and Equation 9 are undetermined. The ρ (in *ZNCC* equation) near zero is also shows the low quality of data for finding the corresponding point. In these cases, the result before calculation is discarded.

6. RESULTS

We stabilized the helicopter sequence by a semi-automatic method (Hoogendoorn *et al.* 2003). We have implemented our algorithms in Matlab. Here we focus on the difficult situations where the tracking by other methods is failed in earlier work (Hoogendoorn *et al.* 2003).

Figure 2 introduces the difficult situations inside our dataset: Similar brightness in the truck image, similar structure in a black car near the road stripe (represented by a red ellipsoid around the black car and a yellow one around the road stripe) and the ambiguity in a car boundary for both black and white car.

Another reason for tracking errors is variation in brightness for both black and white cars. As it is demonstrated in Figure 3, the variation for a specific pixel is very large.

In the above-represented situations, tracking is prone to the wrong results. The results for missing pixel are presented in Figure 4.

In Figure 5, car region, boundary and corner points are extracted in the semi automatic method as described in section 5.1.

The results are improved especially using initial value provided by scale space and boundary constraints. The boundary and region pixels, instead of neighboring pixel of each inside-car pixel calculate the displacement of the car in the successive frames. The results are displayed in Figure 6.

7. CONCLUSIONS

In this paper, we presented the method for the long-term vehicle tracking from aerial video images. We have developed a tracking method based on optical flow using scale space and ZNCC method. The scale space prepared initial value in finding corresponding-stage in the next frame and tracking in the other frames as well. Boundary constraints removed the wrong results.

The experiments show promising results even in very difficult situations such as a dark car in a dark background, small vehicle size, large numbers of vehicles and similar vehicles as well as similar texture.

Using the ZNCC method for corner points, decreases the chance of finding wrong results. Border and region pixels preserve shape as well as brightness. The results also confirmed it. The constant brightness assumption however is not always held for every pixel but for most of them give a correct result. Linear model of brightness is not a correct assumption for every cases but for most of the pixels are correct. However the results of constant brightness assumption are lost after longer frames than linear model of brightness in the similar situation and also using scale space and boundary constraints.

The decision about the best result among redundant results should be constructed based on rigid object assumption which will be presented elsewhere.

ACKNOWLEDGEMENTS

The research presented in this paper is part of the research program "Tracing Congestion Dynamics - with Innovative Traffic Data to a better Theory", sponsored by the Dutch Foundation of Scientific Research MaGW-NWO.

REFERENCES

Atkinson, K., 1996. Close range photogrammetry and machine vision. 12 chapters, 371 pages.

Cohen, I. and Medioni, G., 1999. Detection and tracking moving objects for video surveillance. *IEEE Proc. Computer vision and pattern recognition.*

Coifman, B., Beymer, D., McLauchlan, P., and Malik, J., 1998. A real-time computer vision system for vehicle tracking and traffic surveillance. *Transportation Research: Part C* 6(4), 271–288.

Gorte, B. G. H., Nejadasl, F. K., and Hoogendoorn, S., 2005. Outline extraction of motorway from helicopter image sequence. *CMRT, IAPRS Vienna*.

Harris, C. and Stephens, M., 1988. A combined corner and edge detector. *Proc. 14th Alvey Vision Conf., Univ. Manchester*, 147–151.

Haussecker, H. and Fleet, D., 2001. Computing optical flow with physical models of brightness variation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 661–673.

Hoogendoorn, S. P., van Zuylen, H. J., Schreuder, M., Gorte, B. G. H., and Vosselman, G., 2003. Microscopic traffic data collection by remote sensing. *TRB, Washington D.C.*.

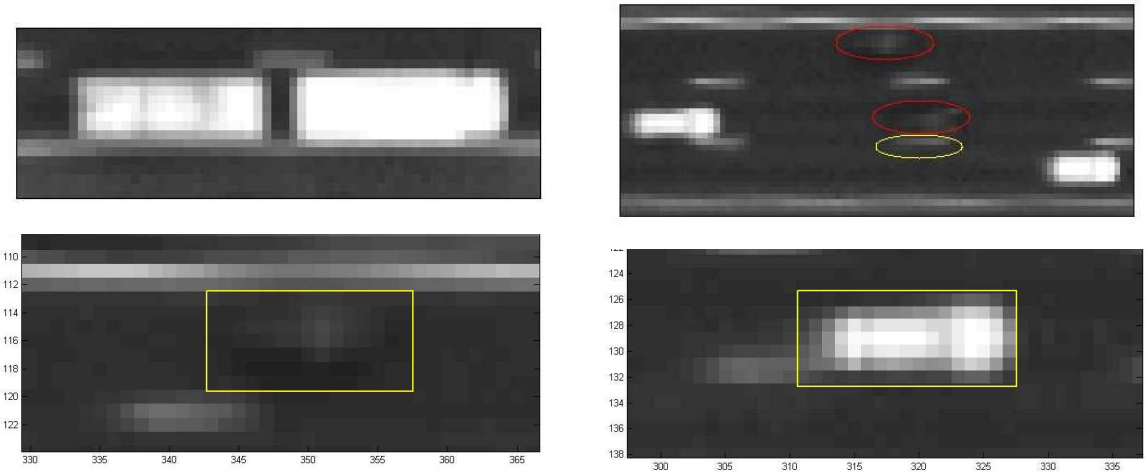

Figure 2. Problems: similar brightness (top left); similar texture (top right); ambiguity in edges (down)

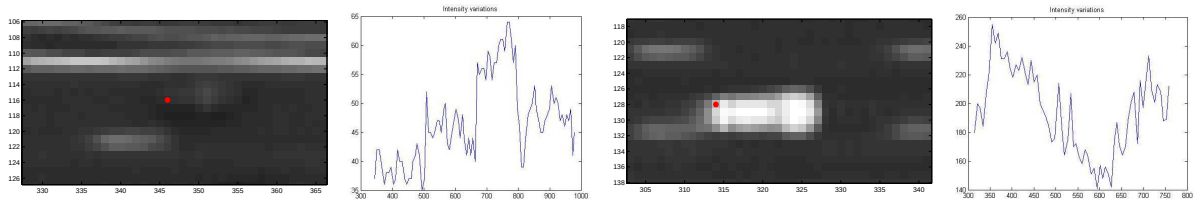

Figure 3. Brightness variations: The position of the specific pixel is shown by red circle (the first image); the brightness variation of this pixel in successive frames is represented as a graph with x-direction is position and y-direction is the brightness (the second image); pixel position, similar for the white car (the third image); brightness variation, similar for white car (the forth image)

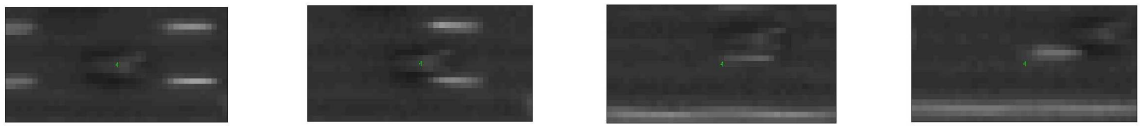

Figure 4. Missing of point tracking: pixel tracking is lost in the third frame

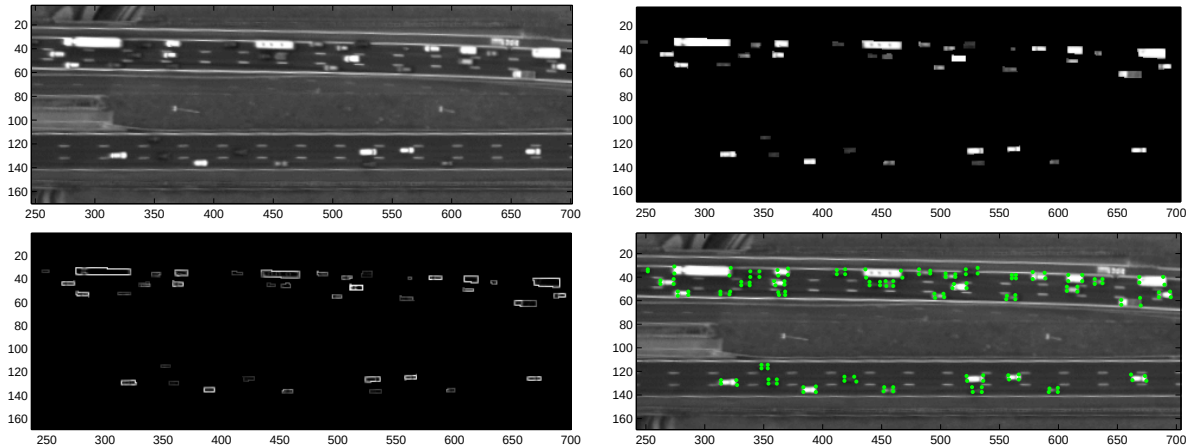

Figure 5. Feature extraction: Image frame (top left); vehicle region extraction (top right); vehicle boundary extraction (down left); corner point extraction (down right)

Hoogendoorn, S. P., Zuylen, H. J. V., Schreuder, M., Gorte, B. G. H., and Vosselman, G., 2003. Traffic data collection form aerial imagery.

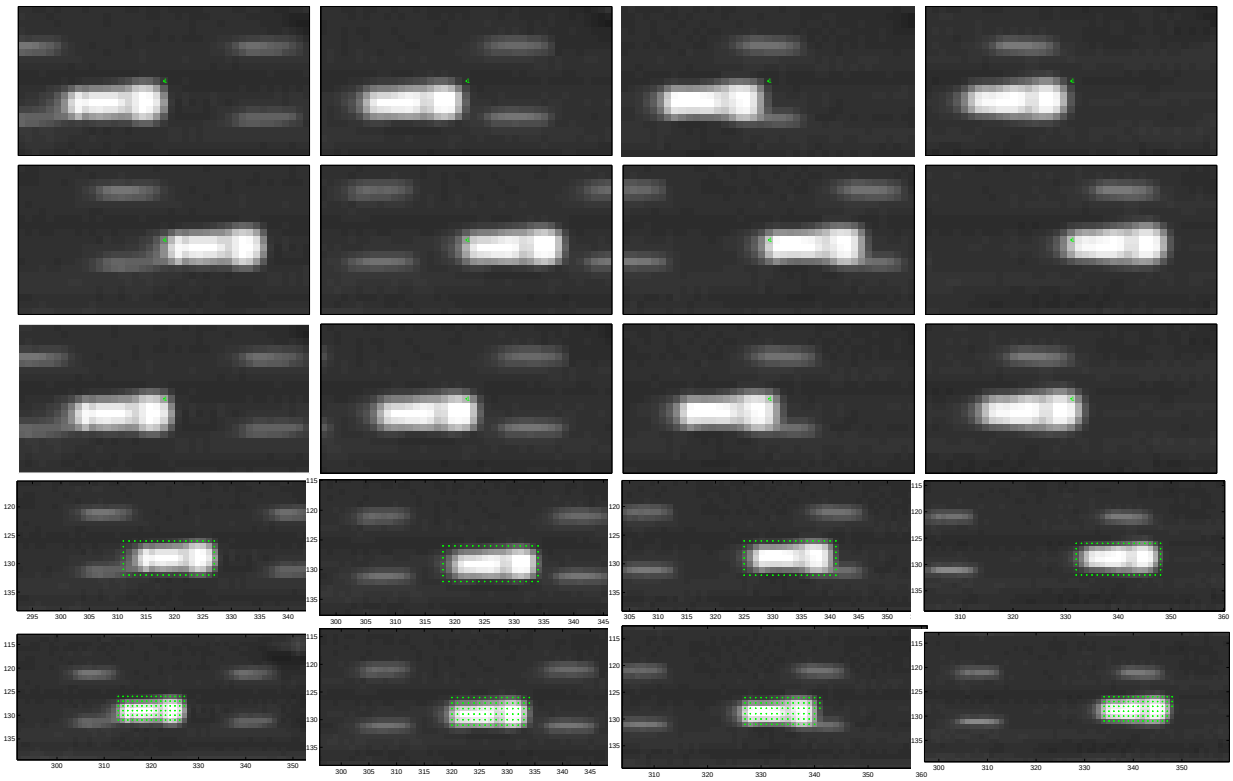

Figure 6. Tracking: the tracking using ZNCC, optical flow assuming constant brightness model, optical flow assuming linear brightness model, optical flow using the boundary pixels, and optical flow using the region pixels is represented respectively in five different rows

IFAC/IFORS.

Jahne, B., 2001. Digital image processing. 6th revised and extended edition.

Lucas, B. and Kanade, T. An iterative image registration technique with an application to stereo vision. *IJCAI81*, 674–679.

Partsinevelos, P., Agouris, P., and Stefanidis, A., 2005. Reconstructing spatiotemporal trajectories from sparse data. *Journal of ISPRS* 60(1), 3–16.

Smith, S. M., 1998. Asset-2: Real-time motion segmentation and object tracking. *Real-Time Imaging* 4(1), 21–40.

Smith, S. M. and Brady, J. M., 1995. Asset-2: Real-time motion segmentation and shape tracking. *IEEE Transactions on pattern analysis and machine intelligence* 17(8), 814–820.

Teunissen, P. J. G., 2003. Adjustment theory, and introduction.