The Statistics of Driving Sequences – and what we can learn from them

Henry Bradler[1], Birthe Anne Wiegand[1]
[1]Visual Sensorics & Information Processing Lab
Goethe University, Frankfurt, Germany
{bradler,wiegand,mester}@vsi.cs.uni-frankfurt.de

Rudolf Mester[1,2]
[2]Computer Vision Laboratory, ISY
Linköping University, Sweden
mester@isy.liu.se

Abstract

The motion of a driving car is highly constrained and we claim that powerful predictors can be built that 'learn' the typical egomotion statistics, and support the typical tasks of feature matching, tracking, and egomotion estimation. We analyze the statistics of the 'ground truth' data given in the KITTI odometry benchmark sequences and confirm that a coordinated turn *motion model, overlaid by moderate vibrations, is a very realistic model. We develop a predictor that is able to significantly reduce the uncertainty about the relative motion when a new image frame comes in. Such predictors can be used to steer the matching process from frame n to frame $n + 1$. We show that they can also be employed to detect outliers in the temporal sequence of egomotion parameters.*

1. Introduction

The motion of typical road vehicles is much more restricted than general motion, as it appears *e.g.* for a handheld camera. A vehicle has a large mass, thus a substantial inertial moment both for translation and for rotation, and even though absolute speed may be very high, the overall motion is very smooth. This smoothness of motion can be expressed in terms of a dynamic model (as this is standard procedure for functions such as ABS, ESP, etc.), and is *predictable* to a very high degree. Predictability and motion constraints are the key to transform the very ill-posed problems of visual motion estimation and structure-from-motion (SFM) into a scheme that is both more efficient and more robust than it can be achieved in the general motion case.

In the following section, we summarize well-known findings from vehicle dynamics [6]; we will later see in section 3 how these facts relate to statistical characteristics of measured car trajectories.

Obviously, most of the translational motion that a car performs is in the forward direction (aligned with the longitudinal symmetry axis of the car) and, depending on steering angle, also slightly sideways. For an idealized car, the rotational motion occurs mostly in the lateral direction (yaw motion, rotation around the vertical axis). Rotation around the longitudinal axis (roll) occurs only as a (deterministic) side effect of sharp turns and pitch motion (rotation around the transversal horizontal axis) is mostly the result of acceleration or braking. We will denote these motions as the nominal motion of the car, and discriminate it from the random exogenous vibrations which are mostly due to deviations of the road surface from a plane. Physical models of car dynamics include the effects of the suspension, of the tire friction, etc. [4]. Here, we will abstract from these effects (even though they might be worthwhile including in further, more elaborated models) and concentrate on a basic vehicle model which is usually denoted as the *coordinated turn model* [1]. This model couples yaw rotation and the change of the translation in the lateral direction.

1.1. The role of vehicle egomotion

The importance of egomotion for efficiently and robustly estimating the motion field cannot be overestimated. Let us regard that point in time when the information from video frame n has been completely processed. As a result of this, we have the 3D egomotion of the car from time $n - 1$ to time n, we have motion vectors (dense or sparse), and we have an estimate of the depth structure of the scene at time n (again: dense or sparse). Obviously, we want to know the next motion from time n to time $n + 1$. Very reasonable estimates for all of these entities can be made by information propagation, that means: given an interpretation of frame n, a very good prediction of frame $n + 1$ can be obtained – even before actually having seen frame $n+1$. Many typical applications, *e.g.* tracking feature points or tracking surface patches, can be stabilized and robustified by making predictions on the basis of the so far only assumed motion from n to $n + 1$.

As any estimate, the estimate of the next motion will be afflicted by errors. These errors can be characterized by covariance matrices. Given these covariance matrices of the motion parameters, the uncertainty reflected in the prediction of *e.g*. point positions or surface element parameters can be computed using nonlinear covariance propagation. From a Bayesian perspective, these covariance matrices represent prior knowledge before a measurement has been made, and the actual measurement, including its own uncertainty, will be fused with the prior using standard Bayesian techniques. We stress here that the result of a measurement taken in the image (*e.g*. a point-to-point match, or a patch-to-patch match) should always be considered as an uncertain result and explicit attention should be directed onto the uncertainty of this measurement.

2. Fundamentals for building statistical egomotion models

In the remainder of this paper, we will first briefly present the theoretical basis for building predictors for time series of random vectors. We will then describe the motion parametrization we use, and point out some characteristics of the KITTI 'ground truth' egomotion sequences. On the basis of measured statistics on that data, a predictor is built which is subsequently evaluated. We conclude with describing an outlier detection scheme based on that predictor, and briefly sketch applications of the predictor in a visual egomotion estimation system.

2.1. Fundamentals of prediction in vector-valued time series

If we examine a sequence of random vectors $\vec{p}$ and intend to build a linear predictor that predicts $\vec{p}[n+1]$ from $\vec{p}[n]$ optimally in the sense of a minimum mean square prediction error, we can apply a well-known result from estimation theory that builds on the assumption that the 1st and 2nd order statistical moments of the random vectors are known [7, p.300].

Let $\vec{p}$ and $\vec{q}$ be random vectors which are distributed according to a multivariate distribution with the mean vector

$$\mathsf{E}\left[\begin{pmatrix}\vec{p}\\ \vec{q}\end{pmatrix}\right] = \begin{pmatrix}\vec{m}_p\\ \vec{m}_q\end{pmatrix} \tag{1}$$

and the joint covariance matrix

$$\mathsf{Cov}\left[\begin{pmatrix}\vec{p}\\ \vec{q}\end{pmatrix}\right] = \begin{pmatrix}\mathbf{C}_{pp} & \mathbf{C}_{pq}\\ \mathbf{C}_{qp} & \mathbf{C}_{qq}\end{pmatrix} \tag{2}$$

Then the conditional distribution of $\vec{p}$, given $\vec{q}$, has the conditional mean

$$\mathsf{E}\left[\vec{p} \mid \vec{q}\right] = \vec{m}_p + \mathbf{C}_{pq} \cdot \mathbf{C}_{qq}^{-1} \cdot (\vec{q} - \vec{m}_q) \tag{3}$$

and covariance matrix of the difference vector $\vec{p} - \vec{q}$

$$\mathsf{Cov}\left[\vec{p} - \vec{q}\right] = \mathbf{C}_{pp} - \mathbf{C}_{pq}\mathbf{C}_{qq}^{-1}\mathbf{C}_{qp}\ . \tag{4}$$

Note that for these relations to hold, it is *not* necessary that any of the involved distributions is Gaussian.

What is presented here is the distribution of a random vector $\vec{p}$ which is statistically dependent on some measurable random vector $\vec{q}$. If the correlation matrices $\mathbf{C}_{pq}$ and $\mathbf{C}_{qq}$ and the mean vectors $\vec{m}_p$ and $\vec{m}_q$ are known, the conditional expectation $\mathsf{E}\left[\vec{p} \mid \vec{q}\right]$ can be determined from the measured vector $\vec{q}$. If additionally $\mathbf{C}_{pp}$ is known, the conditional covariance matrix $\mathsf{Cov}\left[\vec{p} - \vec{q}\right]$ can de determined by using equation 4.

For the application of these general results to the prediction of the pose change, given a temporal series $\{\vec{p}[n]\}$ of pose changes, we just replace $\vec{q}$ by $\vec{p}[n]$ and $\vec{p}$ by $\vec{p}[n+1]$.

2.2. Parametrization of the motion

We parameterize motion as shown in figure 1. We use the camera coordinate system as reference system for the motion calculations, with the camera being mounted on the top of the car as described in [2]. The z-axis coincides with the viewing direction of the camera (optical axis). The x-axis points straight to the right and the y-axis towards the ground.

The angles θ, ψ and ϕ specify the rotation about the x-, y- and z-axis respectively, in mathematically positive sense. This means that θ denotes the pitch, ψ the yaw and ϕ the roll of the camera. It is important to note here that driving straight ahead into the direction of the optical axis of the mounted camera means that the environment is moving towards the camera, resulting in an observation of a movement in negative z direction. To retrieve the direction of the

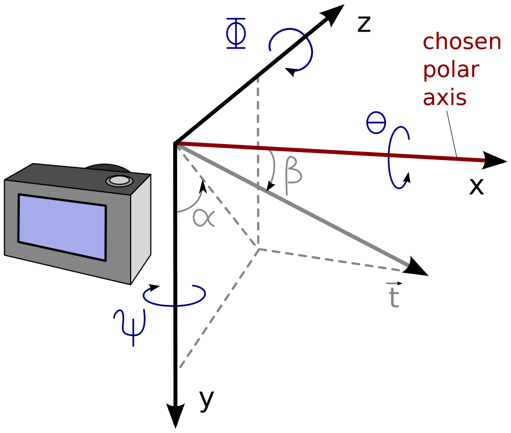

Figure 1: The parametrization for the camera egomotion, showing the camera coordinate system, the 3 rotation angles, and the translation vector $\vec{t}$ which is represented in spherical coordinates by angles α and β, and a length v. In real driving situations, vector $\vec{t}$ is very close to the z-axis ($\alpha \approx \pi/2,\ \beta \approx \pi/2$)

car or camera egomotion, all angles have to be flipped to the opposite direction[1].

Now let $\vec{x}[n]$ and $\vec{x}[n+1]$ be the coordinates of two corresponding points in the camera coordinate system in two consecutive frames. The rigid transformation between these points using a rotation matrix ${}^{n}\mathbf{R}_{n+1}$ and a translation vector ${}^{n}\vec{t}_{n+1}$ is, by convention, defined as follows:

$$\begin{aligned}
\vec{x}[n+1] &= {}^{n}\mathbf{R}_{n+1} \cdot \vec{x}[n] +^{n} \vec{t}_{n+1}, && (5)\\
{}^{n}\vec{t}_{n+1} &= v \cdot \begin{pmatrix} \cos\beta \\ \sin\beta\cos\alpha \\ \sin\beta\sin\alpha \end{pmatrix} && (6)\\
{}^{n}\mathbf{R}_{n+1} &= \mathbf{R}_z(\phi)\cdot\mathbf{R}_y(\psi)\cdot\mathbf{R}_x(\theta) && (7)\\
\mathbf{R}_x(\theta) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} && (8)\\
\mathbf{R}_y(\psi) &= \begin{pmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{pmatrix} && (9)\\
\mathbf{R}_z(\phi) &= \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} && (10)
\end{aligned}$$

The angles α and β describe the direction of the translation vector in spherical coordinates. β is the polar angle relative to the x-axis (our polar axis), and α is the azimuth angle measured relative to the y-axis in the y-z-plane. We chose this constellation in order to obtain an operation point of the parameters which is far from a 'gimbal lock' situation when moving roughly along the optical axis. For cameras looking into strongly different directions (*e.g.* sideways), the parametrization has to be adapted, of course.

A movement that perfectly follows the optical axis results in $\alpha = -\pi/2$ and $\beta = \pi/2$, as the environment moves directly towards the viewer when moving forward.

These parameters are stacked into a combined (relative) motion parameter vector as follows:

$$\vec{p} \overset{def}{=} \begin{pmatrix} \text{pitch } \theta \\ \text{yaw } \psi \\ \text{roll } \phi \\ \text{translation length } v \\ \text{spherical angle } \alpha \\ \text{spherical angle } \beta \end{pmatrix} \qquad (11)$$

We explicitly separate the *length* of the translation vector $\|\vec{t}\| = v$, measured in meters, from the translation *direction*, since this entity is not *observable* (in the control theory sense [1, 4]) for a monocular camera. If we know the frame rate, we can easily determine the velocity. In order to support intuitive interpretation of numerical values, we convert this into km/h when applicable.

The conversion from v and the spherical coordinates of the translation, α and β, to Euclidean coordinates and vice versa is straightforward. The separation between direction and translation length also allows to see the correlations between certain parameters that would otherwise not be as easily identifiable: The yaw angle ψ correlates with β, the spherical angle caused by horizontal displacement of the translation. We expect a clear correlation between these two parameters. Weaker, but still noticeable correlations should be visible between the two aforementioned entities and the integral of the roll angle ϕ (much better visible in a car with an elevated center of mass), as well as between the acceleration (the derivative of the velocity), the pitch angle θ and α, the spherical angle that denotes the vertical portion of the movement.

An additional benefit of this parameterization exists when applying it to a monocular setup, where the scale cannot be determined: As v is the only dimension influenced by the scale at all, and fully determined by it, it can just be ignored as long as no depth information is available, resulting in a five dimensional parameterization.

The only problem that spherical angles produce is their uncertainty for very slow motion. We observed unstable behavior in the ground truth data for velocities below approximately 3 km/h and therefore had to include vehicle halt detection to stabilize them, as described next.

2.3. Data problems in original KITTI ground truth

While working on the KITTI sequences [2, 3] and comparing image data and egomotion ground truth data coming with KITTI, we occasionally discovered strange and unexpected behavior of the ground truth parameters θ, ψ, ϕ, v, α and β. We identified the following categories:

- vehicle halt (cf. figure 2)
- 'IMU freeze' (cf. figure 3)
- oscillation (cf. figure 4)

The first category (vehicle halt) does not necessarily lead to problematic situations, but for our spherical representation of the translation it does. When the car is driving very slowly ($<$ 3 km/h), the direction of the inter-frame translation (velocity) is very hard to estimate correctly. To prevent this effect from influencing our statistic model, we *correct* the KITTI ground truth by manually setting the direction of the translation to a straight forward direction ($\alpha \overset{!}{=} -\pi/2,\ \beta \overset{!}{=} \pi/2$)[2] when velocities below 3 km/h are observed. The effect of situations of this type on the motion parameters can be seen in figure 2.

[1] $(\theta, \psi, \phi, \alpha, \beta) \rightarrow (-\theta, -\psi, -\phi, \alpha + \pi, \pi - \beta)$

[2] We emphasize that driving straight ahead into the direction of the optical axis of the mounted camera means that the environment is moving towards the camera, thus into the opposite direction of the optical axis. This direction is encoded by $(\alpha,\ \beta) = (-\pi/2,\ \pi/2)$ in our parameterization.

The second category occurs only in the first sequence (KITTI sequence 0) and looks like a linear interpolation between two time steps when the IMU was unable to deliver ground truth for the values in between (cf. figure 3).

We were also able to observe a strange behavior of the KITTI ground truth when the car is making a turn. The third category (oscillation) is characterized by a noticeable oscillation of the velocity v and the horizontal direction of the translation β. This behavior can not be explained by looking at the images: The car is driving smoothly and we would expect smoothly varying motion parameters as well. The effect on the motion parameters can be seen in figure 4.

3. Setting up the dynamic model

In this section we will construct a linear predictor for the motion parameters based on the statistics of the motion ground truth provided by KITTI and by doing so, we will also analyze auto-correlation and temporal cross-correlation of the motion parameters. The goal is to find the best prediction of the motion parameters for the next image ($\vec{p}[n+1]$), given the values for the current image ($\vec{p}[n]$). It would of course be possible to use a higher order model which could then account for the *change* of the parameters, *e.g.* the acceleration, but for the course of this paper, we focus on the first order model.

3.1. Learning the motion statistics

In section 3.2, we will use equation 3 to build a one-step linear predictor for the motion parameter vector $\vec{p}[n+1]$ given the previous one: $\vec{p}[n]$

$$\hat{\vec{p}}[n+1] \overset{def}{=} \mathsf{E}\left[\vec{p}[n+1] \mid \vec{p}[n]\right]. \tag{12}$$

In order to evaluate this expression, we need to measure/estimate the motion parameter mean $\vec{m}_p$, $\vec{m}_q$ and cross- and autocorrelation $\mathbf{C}_{n+1,n}$, $\mathbf{C}_{n,n}$. In this case $\vec{p}$ and $\vec{q}$ are the same observation (motion parameters). The only difference is that $\vec{q}$ is delayed by one time step ($\Delta n = 1$) with respect to $\vec{p}$. We calculated these statistical moments on a set of all motion parameters of KITTI sequences 0 - 10. This results in a statistic based on more than 23,000 motion parameter vectors that cover a wide range of driving scenarios (city, highway, rural road, etc.), which gives us the advantage of being able to treat $\vec{p}[n]$ as a stationary process (*i.e.* its statistics are independent of the current frame number n). For the KITTI ground truth we obtained

$$\vec{m}_p = \begin{pmatrix} 5.554\text{ e-05} \\ -4.428\text{ e-04} \\ -1.732\text{ e-06} \\ 0.000\text{ e+00} \\ -1.551\text{ e+00} \\ 1.569\text{ e+00} \end{pmatrix}, \tag{13}$$

where we forced the mean of the velocity (4th component) to be zero[3]. We did so because any stable predictor must exhibit a trend to the mean (beside the correlation to the old value), and in a realistic setting this trend should be to decreasing velocity due to friction, but not to the random mean speed observed in the KITTI sequences. We present the autocovariance matrix $\mathbf{C}_{n,n}$, here in *normalized* form, that is, in terms of a variance vector $\mathsf{Var}\left[\vec{p}\right]$ and a normalized covariance matrix $\overline{\mathbf{C}}_{n,n}$:

$$\overline{\mathbf{C}}_{n,n} = \begin{pmatrix} 1.00 & 0.08 & 0.03 & 0.01 & 0.04 & -0.05 \\ 0.08 & 1.00 & 0.13 & -0.04 & 0.03 & \text{-0.60} \\ 0.03 & 0.13 & 1.00 & -0.01 & 0.01 & -0.27 \\ 0.01 & -0.04 & -0.01 & 1.00 & -0.05 & 0.05 \\ 0.04 & 0.03 & 0.01 & -0.05 & 1.00 & -0.06 \\ -0.05 & \text{-0.60} & -0.27 & 0.05 & -0.06 & 1.00 \end{pmatrix}, \tag{14}$$

$$\mathsf{Var}\left[\vec{p}\right] = \begin{pmatrix} 9.118\text{ e-06} \\ 2.989\text{ e-04} \\ 6.935\text{ e-06} \\ 1.896\text{ e-01} \\ 2.535\text{ e-03} \\ 1.362\text{ e-03} \end{pmatrix}. \tag{15}$$

The only strong intra-vector correlation which is visible at first glance is the one between the yaw angle ψ and the horizontal angle of translation β. In addition to that, but much weaker, there is also a correlation between the roll angle ϕ and the mentioned two entities.

The cross-covariance $\mathbf{C}_{n+1,n}$ is represented by the *normalized* covariance matrix $\overline{\mathbf{C}}_{n+1,n}$

$$\overline{\mathbf{C}}_{n+1,n} = \begin{pmatrix} 0.61 & 0.08 & 0 & 0.01 & -0.01 & -0.05 \\ 0.08 & 1.00 & 0.12 & -0.04 & 0.03 & 0.60 \\ 0.01 & 0.15 & 0.64 & -0.01 & 0 & -0.22 \\ 0.01 & -0.04 & -0.01 & 1.00 & -0.05 & 0.05 \\ 0.05 & 0.03 & 0 & -0.05 & 0.98 & -0.05 \\ -0.04 & \text{-0.60} & -0.23 & 0.05 & -0.06 & 0.78 \end{pmatrix}, \tag{16}$$

and the cross-covariance vector $\text{diag}(\mathbf{C}_{n+1,n})$

$$\text{diag}(\mathbf{C}_{n+1,n}) = \begin{pmatrix} 5.522\text{ e-06} \\ 2.978\text{ e-04} \\ 4.423\text{ e-06} \\ 1.894\text{ e-01} \\ 2.489\text{ e-03} \\ 1.068\text{ e-03} \end{pmatrix}. \tag{17}$$

These statistics show that during normal driving the yaw angle ψ and the velocity v (length of translation vector) only change very slowly, which leads to an excellent predictability from their temporally preceding values. The significant non-zero off-diagonal elements reflect the restricted motion

[3] Cf. definition 11 for the kinematic meaning of the vector elements.

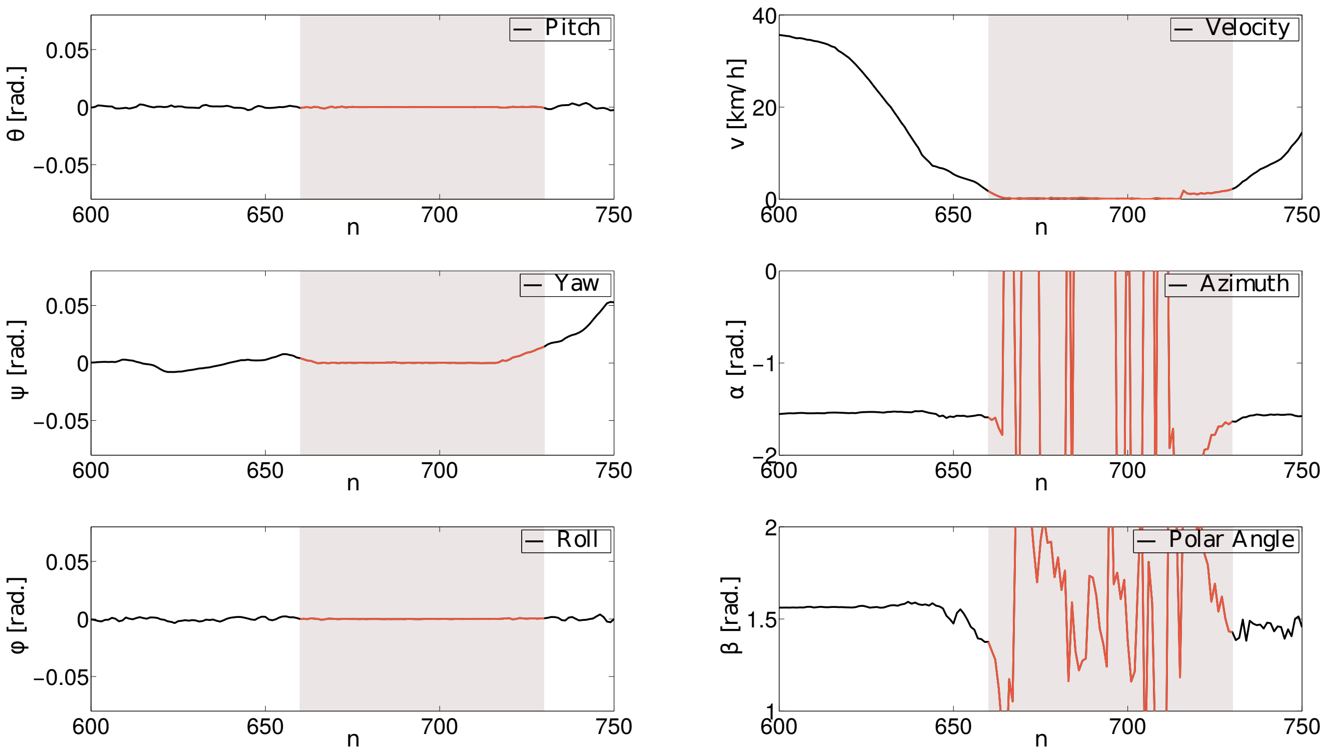

Figure 2: Typical behavior of motion parameters during a vehicle halt (red line). This example is from KITTI sequence 7, frames $660 - 730$, but similar effects can also be observed *e.g.* in sequence 0, frames $540 - 560$ and sequence 5, frames $2330 - 2400$.

of a car, which can very well be described by a *coordinated turn model* [1, 4], and the influence of a lateral turn on the roll angle.

3.2. The one-step linear predictor

Now we can finally instantiate the predictor matrix $\mathbf{A}$ according to

$$\mathbf{A} \overset{def}{=} \mathbf{C}_{n+1,n} \cdot \mathbf{C}_{n,n}^{-1} \tag{18}$$

which defines the predictor as

$$\hat{\vec{p}}[n+1] = \mathbf{A} \cdot (\vec{p}[n] - \vec{m}_p) + \vec{m}_p, \tag{19}$$

with

$$\mathbf{A} = \begin{pmatrix} 0.60 & 0.01 & -0.03 & 0 & 0 & 0 \\ 0.01 & 0.99 & -0.12 & 0 & 0 & 0 \\ -0.01 & 0.01 & 0.63 & 0 & 0 & 0 \\ -0.40 & 0.05 & 0.04 & 1.00 & -0.02 & 0.01 \\ 0.21 & 0 & -0.10 & 0 & 0.98 & 0 \\ 0.06 & -0.43 & -0.27 & 0 & -0.01 & 0.66 \end{pmatrix}. \tag{20}$$

4. Experiments

4.1. Experimental evaluation of the predictor

Before trying the predictor on the KITTI odometry data, we have a look at the theoretical gain obtained from the predictor. Table 1 lists the variances of the motion parameters and the residuals of their one-step estimates vs. ground truth for KITTI sequences 0 – 10. From these results we obtain the following insights:

- The motion parameters of driving scenes are strongly constrained in terms of the value range that typically appears in the vast majority of driving situations. Their first order distribution is described by statistical 1st and 2nd order moments (equation 13 and 14).

- For typical driving scenarios, a linear predictor like given in equation 18 – 20 provides a very substantial reduction of motion uncertainty *before* the next image even has been acquired. The parameters yaw ψ, translation length v, and spherical angle α are particularly well predictable.

- Simply taking the previous motion as a predictor for

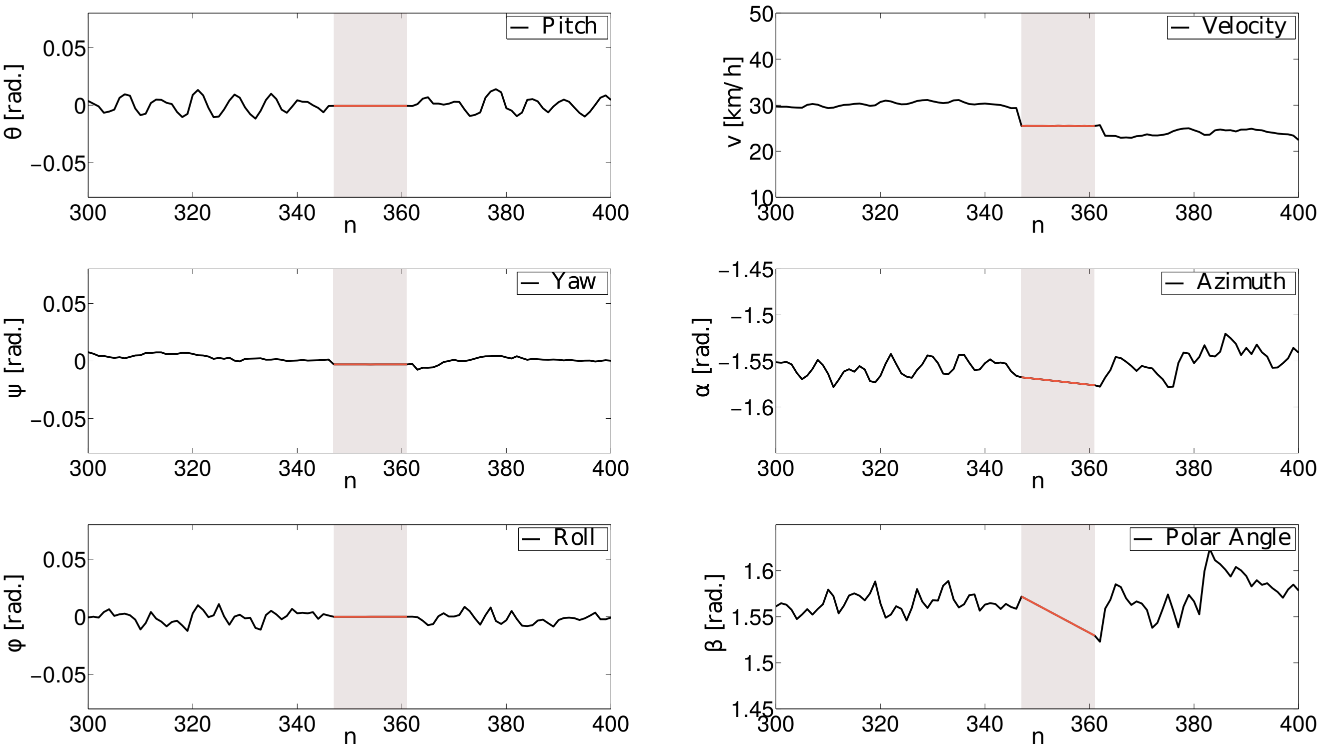

Figure 3: Behavior of motion parameters during 'IMU freeze' (red line). Example taken from KITTI sequence 0, frames $345-360$. Similar effects can also be observed in *e.g.* frames $1915-1930$ and frames $2275-2290$ of the same sequence.

$\vec{p}$	Var $[\vec{p}]$	Var $\left[\hat{\vec{p}}-\vec{p}\right]$	ratio
θ	9.118e-06	5.751e-06	0.631
ψ	2.989e-04	2.224e-06	0.007
ϕ	6.935e-06	4.081e-06	0.588
v	1.896e-01	4.518e-04	0.002
α	2.535e-03	9.076e-05	0.036
β	1.362e-03	4.885e-04	0.359

Table 1: Variances of motion parameters without prediction (col.1), predictor residuals (col 2), and the resulting variance ratio (col 3). Small ratios → good predictability.

the next motion is not bad, but inferior to a trained predictor as presented here.

4.2. Using the predictor to flag outliers in egomotion

Plotting the predictor output vs. the true motion parameters does not show any significant visual difference, except in those situations when the actual value deviates strongly from the prediction. This property can be used to flag egomotion estimation errors both in the ground truth as well as when egomotion is computed from visual data.

The recipe for outlier detection is as simple as it is effective: From the predictor equations, we obtain the covariance matrix of the predictor residual. Each time a new motion vector has been computed (or read in from ground truth data), it is compared against the predicted value, and the resulting *residual* is processed using the residual covariance matrix as a metric. In other words, we compute the Mahalanobis distance of the regarded motion vector w.r.t. the predicted motion value. If this value is too high, the motion vector is considered as questionable.

4.3. Using statistics from state-of-the-art egomotion estimators

We have seen that the KITTI ground truth does actually contain occasional outliers. It remains to be seen whether this has a significant influence on the learnt statistics, and subsequently also on the predictor and the capability to exploit the predictor as a detector for probable outliers.

When going beyond using the KITTI egomotion ground truth data, and using egomotion data from a good state of the art method, we see that smooth and realistic behavior can also be achieved for the horizontal part of the translation vector, angle β in our parametrization. Performing the same statistical analysis on such egomotion data leads to stronger correlation between ψ, ϕ and β and a much stronger tempo-

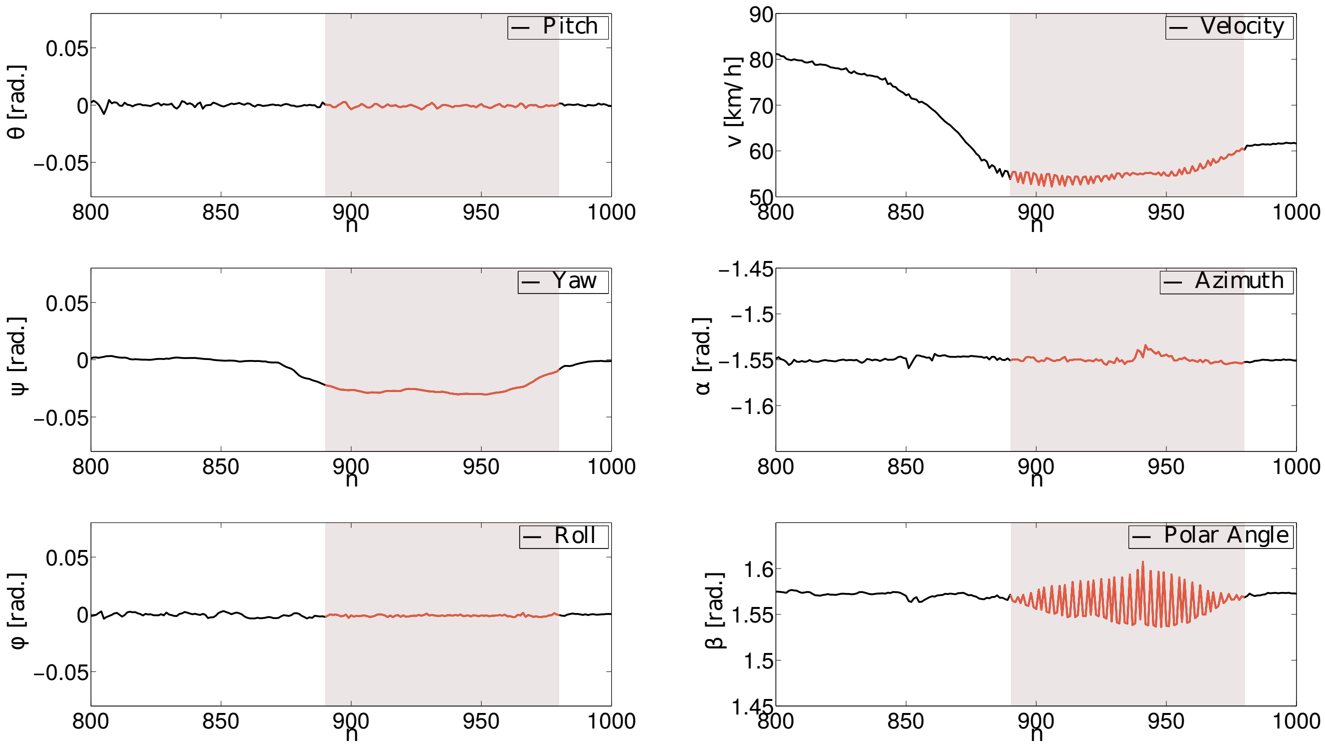

Figure 4: Oscillating behavior of motion parameters while cornering (red line). This example is taken from KITTI sequence 1 frames 890−980 but can also be observed i.a. in sequence 6 frames 280−320 and 680−720 and sequence 10 frame 850−880.

ral correlation and thus predictability.

In order to evaluate which results can be expected if a state of the art method is used for visual egomotion estimation, we considered the work of Persson *et al.* [5]. The authors of this paper kindly provided the pose data computed with their method such that we can directly compare against KITTI ground truth. Both curves are shown in figure 5. Additionally, we felt free to also overlay the results from a very recent, yet unpublished, method from our own lab drawn in the same figure ('OurOwn').

In figure 5 we see that significantly less spurious oscillations occur for the employed methods for visual egomotion estimation. This leads to more realistic statistics, which in return allows for a much better detection of outliers. This even works if statistics are calculated on the motion parameters of these methods and then used to detect outliers in the motion parameters of other methods (e.g. KITTI), cf. figure 6.

These results indicate that it is possible to reduce outliers even below the level provided by [5], which is today (Sep 2015) ranking high on the KITTI odometry benchmark. We used the statistics learnt on the [5] pose data to train a predictor, and applied it to the KITTI ground truth data. We used the error covariance matrix from this new predictor as a metric to rate the (vector-valued) residuals obtained for this predictor. Figure 6 shows the temporal course of the motion data for a critical section of the drive. The lower left graph shows the course of the Mahalanobis distance based on the predictor trained on KITTI ground truth, whereas the lower right graph shows the Mahalanobis distance when using the statistics learnt from the [5] pose data. The conclusion is that advanced methods like [5] provide 'cleaner' statistics for training an egomotion predictor.

5. Conclusions

We have shown how to build a frame-to-frame predictor for the egomotion of a typical road vehicle on the basis of ground truth data obtained (in the present case) from the renowned KITTI benchmark data base. We showed that typical characteristics of vehicle motion are reflected in the statistics of egomotion data, and also represented in predictors which are based on these statistics. The presented scheme can be used to monitor the credibility of individual egomotion estimates, and also can be employed to steer matching and tracking approaches.

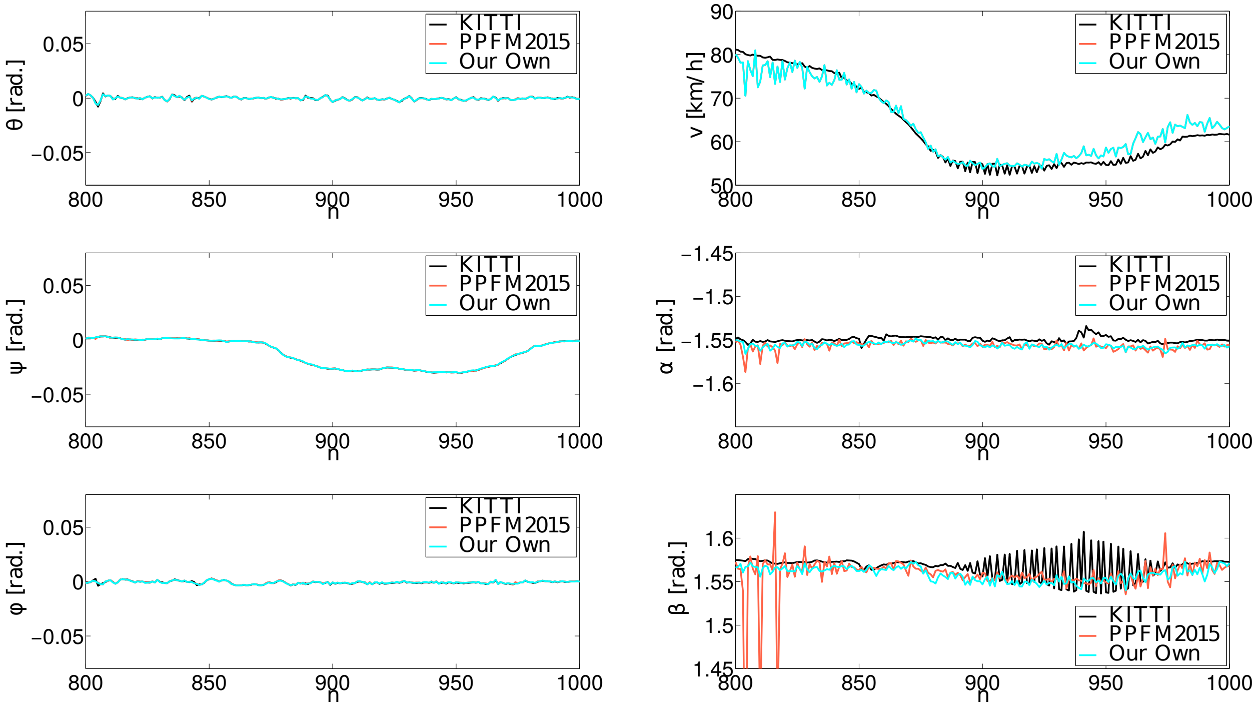

Figure 5: Comparison of the motion parameters provided by KITTI ground truth, versus *Persson* et al. *2015* (PPFM2015) and a very new method from our own group ('Our Own'), which has not been published yet. Best viewed in color.

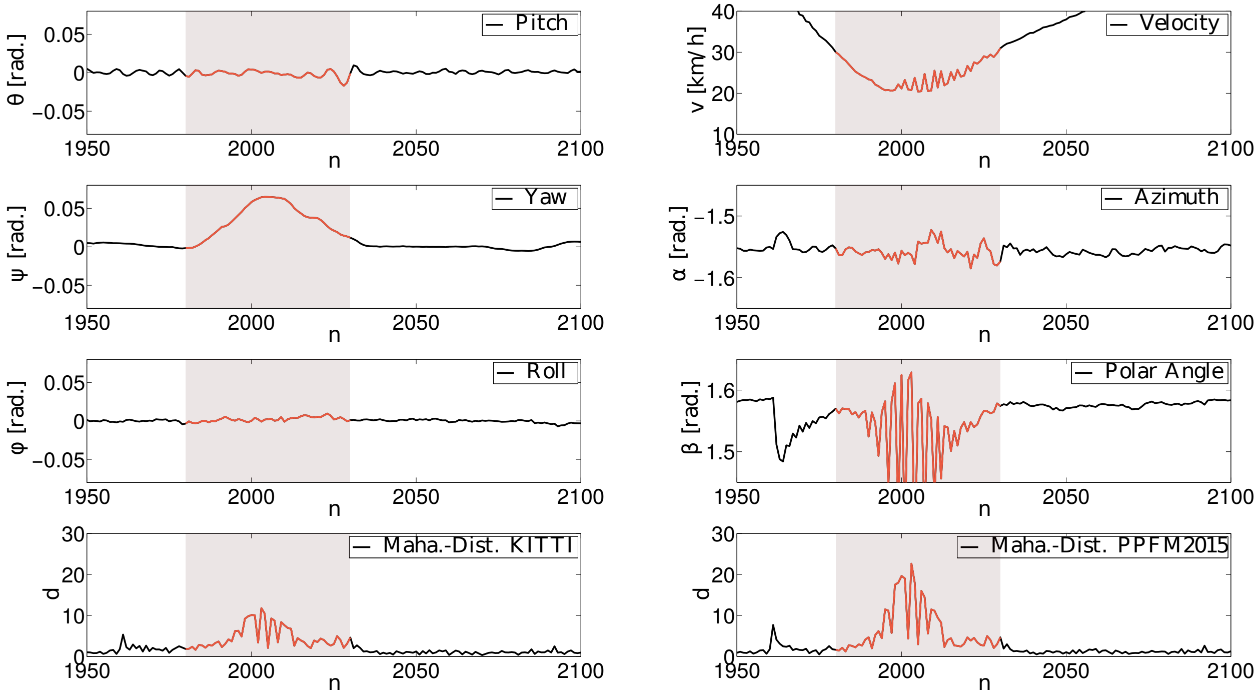

Figure 6: Motion parameter sequence (top 3 rows) and Mahalanobis-weighted predictor residuals (4th row) for a critical section from the KITTI sequences. The detection of unreliable sections (marked in red) based on the predictor residuals uses the statistics from KITTI ground truth (lower left) and from [5] (lower right).

References

[1] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan. *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004. 1, 3, 5

[2] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, page 0278364913491297, 2013. 2, 3

[3] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the KITTI vision benchmark suite. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 3354–3361, 2012. 3

[4] F. Gustafsson. *Statistical sensor fusion*. Studentlitteratur. Professional Pub Serv, 2010. 1, 3, 5

[5] M. Persson, T. Piccini, M. Felsberg, and R. Mester. Robust stereo visual odometry from monocular techniques. In *Proc. Intelligent Vehicles Symposium*, Seoul, 2015. IEEE. 7, 8

[6] G. Rill. *Road Vehicle Dynamics: Fundamentals and Modeling*. CRC Press, September 2011. 1

[7] L. L. Scharf. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison-Wesley, New York, 1991. 2