On Feature Templates for Particle Filter Based Lane Detection

Andre Linarth and Elli Angelopoulou

Abstract—In this work we propose the application of state-of-the-art feature descriptors into a Particle Filter framework for the lane detection task. The key idea lies on the comparison of image features extracted from the actual measurement with a priori calculated descriptors. First, we demonstrate how a feature expectation can be extracted based on a particle hypothesis. We then propose to define the likelihood function in terms of the distance between the expected feature and the features calculated from the current measurement. We select the Histogram of Oriented Gradients as a descriptor and the Battacharyya distance as a metric. We show that this simple approach is powerful in terms of pattern discrimination and that it opens a new set of possibilities for increasing the robustness of lane detectors.

I. INTRODUCTION

In many Driver Assistance Systems, the analysis of vehicle surroundings plays an important role. Among all the elements around the car, the driving lanes can be seen as a base component for applications like Lane Departure Warning, Lane Keeping Assistance, Curve Speed Warning, and others. For applications involving obstacle avoidance, as for example Adaptive Cruise Control or Collision Warning System, lane information is essential in supporting the assessment of the relevance of targets and determining the associated risk. For these reasons, a robust lane detection system is necessary. Approaches to detect lanes from video images dates back to the 1980s. However, due to an extensive variation of environmental conditions and variability of road types, there are still several challenges to be solved or alleviated. A wide range of algorithms has been proposed so far, applying different types of sensors, models, features, estimation and tracking techniques. The reader is referred e.g. to [15] and [11] for reviews on lane estimation systems. Since the work of Southall and Taylor [16] and Apostoloff and Zelinsky [1], the use of Particle Filters as a framework for detection and tracking of lane boundaries gained significant momentum. These filters offer a number of advantages. They are based on a probabilistic framework and can handle arbitrary distribution types. Additionally, incorporating new cues to enhance the robustness of the estimation is straightforward. Recently, Zhou et al. [17] proposed a maximum a posteriori formulation for the lane detection, applying a Particle Filter framework to simultaneously track lane shape and vehicle position. Franke et al. [7] applied such a framework to the recognition of country roads, and Loose et al. [13] for rural roads. In [10], Kim mentions that due to vehicle vibration, tracking the motion of lane boundaries is more appropriately modeled with a Particle Filter rather than a Kalman Filter. In [12], Liu et al. propose a combination of statistical Hough transform to Particle Filters. More recently, in [6], Danescu e Nedevschi studied the appliance of a complex lane model (11 dimensions) as well as stereo vision cues into a Particle Filter based lane tracking.

In contrast to previous approaches, we concentrate on a new definition for the likelihood estimation process. The main contribution of our approach is to apply state-of-the-art feature descriptors and distance metrics directly to the weighting process of a particle. The presentation follows the typical steps of Particle Filter design. First, our parameter space is defined in Section II-A, consisting of our lane model. Then, we concentrate on the derivation of the likelihood model, where the concept of feature templates is introduced. The evaluation of the posterior distribution is discussed subsequently. We conclude the description by defining resampling and update processes.

II. METHOD

In this work we perform lane detection based on a single monocular camera. Images are undistorted and warped to a virtual view where the ground plane is parallel to the image plane (also called Inverse Perspective Mapping [14], or Bird Eye View). Camera parameters are derived from an offline calibration process. The lane detection algorithm operates on a Particle Filter [9] framework. Compared to Kalman Filters, Particle Filters are not bounded to Gaussian distributions, and represent a more appropriate framework for the lane detection task, which essentially is a multi-modal problem. For instance, in the case of road splitting and merging situations, multi-lanes, or the presence of mismatching patterns, more than two lanes can be observed. We base our method on the well established Sample Importance Resampling algorithm [9], with adaptations on the resampling strategy. For the final estimates, we segment the posterior distribution using the Mean-Shift [4] algorithm and select the distribution modes as output models.

A. Lane Model

Our lane model is composed of two second order curves representing the left and right lane boundaries. The curves are defined similarly to the widely applied Clothoid models. However, we discard the component allowing the variable curvature. Thus,

\[ x^L(d) = d \]  \hspace{1cm} (1)
\[ y_{l,r}^L(d) = \frac{c_{l,r} \cdot d^2}{2} \pm \frac{w}{2}, \]  

(2)

where \( x(d) \) and \( y_{l,r}(d) \) are an approximation of the Cartesian coordinates of the point at a distance \( d \) along the curve. The indexes \( l \) and \( r \) indicate left and right lane boundaries, respectively. \( c_{l,r} \) is the curvature parameter and \( w \) the lane width. The superscript \( L \) denotes the lane coordinate system, defined on the road plane. Mapping to the vehicle coordinate system (superscript \( V \)) is achieved through an in-plane rotation and a translation with respect to the \( y \)-axis, thus

\[
\begin{pmatrix}
  x^V_l \\
  y^V_{l,r}
\end{pmatrix}
= \begin{pmatrix}
  \cos \alpha & -\sin \alpha \\
  \sin \alpha & \cos \alpha
\end{pmatrix}
\begin{pmatrix}
  0 \\
  y_{off}
\end{pmatrix}
\begin{pmatrix}
  x^L_l \\
  y^L_{l,r}
\end{pmatrix},
\]  

(3)

where \( \alpha \) is the yaw angle and \( y_{off} \) represents the lateral offset of the lane with respect to the vehicle. The vehicle coordinate system is defined also on the road plane with its \( x \)-axis aligned at the vehicle center in the driving direction. This model offers the advantage of relaxing the assumption of parallel markers, while keeping the parameter space relatively small (5 degrees of freedom). In fact, lanes are forced to be parallel only in their origin, once both parabolas share the same rotation angle. The parameters of equations (1) to (3) define the \( i^{th} \) particle vector:

\[ p^i = [y^V_{off}, \alpha^i, w^i, c_l^i, c_r^i] \]  

(4)

B. Measurement Model

In this work, a probabilistic framework is used to derive the current lane model. The goal is to estimate a Probability Density Function (PDF) on the parameter space, and based on that, to derive the actual lane parameters. In a Particle Filter framework, this PDF is estimated by a set of discrete samples weighted proportionally to their probability of lying on a lane (see e.g. [2] and [3] for details). The definition of the sample weighting is a key step of a Particle Filter based system. The main component of this definition is the so called Likelihood function, which depicts the probability a measurement describes a given state. In this paper we propose a simple, yet powerful idea to apply state-of-the-art computer vision descriptors and metrics in the definition of this Likelihood function. The key insight is that, once the state to be evaluated is given in a Particle Filter, it is also possible to derive an expectation of how an image feature should look like given this state. Therefore, a Likelihood function can be derived by comparing expected and measured features through some distance metric. Besides its simplicity, one of the main advantages of such an approach is that measurement descriptors can be shared while evaluating multiple descriptors (or patterns, respectively). Sharing can also occur among applications, for example in the classification of further road surface markings.

1) Grid and Window Selection: For efficiency purposes, measurements are taken on a discrete grid in image space. Every feature descriptor is computed over a sliding window which is composed of multiple grid cells (see Fig. 1). Image borders are excluded from the window calculation. This grid and window structure together with the particle hypothesis are essential in relating templates and measurements. This is done as follows. Each particle hypothesis is expressed in terms of the vehicle coordinate system using equations (1) to (3) and then further projected to the image space. The intersection of such hypothesis with the horizontal grid borders is defined as a key point. Then, given our fixed window size (i.e. here more specifically the height of the window), we obtain several possible entry-exit key point pairs. Each entry-exit pair defines a line and thus a corresponding template. Since windows overlap, we select as measurement window the one in which the entry key point is located closest to its center. As we expect only small slopes, the target pattern will therefore be approximately centered in the window. This is illustrated in Figure 1.

2) Template Generation: The process of generating feature templates is executed offline. By assuming the orthogonal projection of the ground plane resulting from the Bird’s Eye View, only a small subset of rotation and translation combinations of the marker inside a window needs to have its feature descriptors calculated. These transformations represent however a continuous space, thus necessitating of discretization. For simplicity, we round (to an integer representation) the intersection points between window boundaries and the hypothesized model. Centered at these points, a template image of a simple white marker on a white background is generated. The points are first projected into the road coordinate system, and marker patterns of a given width (in our case, 15cm or 30cm) are calculated. The resulting marker is back projected into image space, where finally the drawing into the template image is performed. From this image, the features are calculated. Examples for template images are shown in Figure 2.
3) Feature Descriptor: Histograms of Oriented Gradient (HoG) [5] were selected as feature descriptors, due to their shape description properties and their robustness to varying illumination conditions. Lane markers are expected to exhibit higher brightness than the surrounding road. Histogram bins are spread in the entire $360^\circ$ range. The selection of the cell size is therefore related to the image resolution and marker size. Recall that the descriptor drops spatial information of edge positions inside a cell. This aspect is important for lane detection, as the distance between edges is a very discriminative feature. To ensure sufficient spatial discrimination of the lane edges, a smaller cell width is preferred. To reduce the amount of histograms in the descriptor, we omit block normalization and instead normalize over the entire window.

4) Distance Metric: Histograms are discrete representations of a distribution. Thus, in principle, any metric measuring distances between distributions could be applied in this context. However, by the nature of our problem, especially with the presence of synthetic templates, a number of these histograms contain only zero entries in their bins. This is an expected effect, which indicates that the road region contains no texture. However, many metrics (e.g. Bhattacharyya, Kullback-Leibler divergence and others) are not defined when one of the distributions is uniformly zero. Instead of simply selecting a metric that is resilient to such problems (e.g. the L2 norm), or to circumvent each case individually (e.g. by applying small constants in the zero entries case), we rather prefer to make our templates (a little) more similar to what it is expected in real images. This is achieved by applying a minimal amount of noise to each synthetic image, causing the descriptor to reflect it in its entries. We iterate this process several times, and select the average of the descriptors as our feature template. As distance metric, we empirically chose the Bhattacharyya distance:

$$
\zeta(h^1, h^2) = \frac{1 - \sum_{b \in B} \sqrt{h^1_b h^2_b}}{\sum_{b \in B} h^1_b \cdot \sum_{b \in B} h^2_b},
$$

which compares the distance between two normalized histograms $h^1$ and $h^2$. $B$ denotes the number of bins in the histogram, and $h_b$ represents the value stored in its $b$-th bin. From this distance we define the $i$-th particle weight $\eta^i_{l,r}$ of the left $l$ or right $r$ boundaries as

$$
\eta^i_{l,r} = \sum_N \sum_H e^{-\frac{\zeta(h^{m,1}, h^{1,1})}{2\lambda_m}},
$$

where $\lambda_m$ is a parameter representing the degree of confidence of the measurement term. $N$ is the number of evaluated windows, as selected through the intersection of the particle with the measuring grid, $H$ is the number of histograms contained in one feature vector. The terms $h^{m,1}$ and $h^{1,1}$ are one of the histograms associated with respectively, the measurement and the template for the given window and particle hypothesis. Note that no normalization is applied with respect to the number of windows, therefore particles whose projections lie on the image area are favored. While this can be seen as a bias in the estimation, it is an important factor in penalizing particles that diverge from the measurement boundaries (e.g. a particle with only one edge hitting the image).

![Fig. 3. Two examples of distance metric response for vertical markers. (a) and (c) are the Bird’s Eye View of the captured images. (b) and (d) are the respective likelihood estimates (normalized for visualization purposes).](image)

Figure 3 shows typical responses of the distance metric, when applied with a single, vertical, 30cm wide marker template into the entire measurement grid. In the images, it is possible to observe the discriminating effect (or filtering performance) achieved through this approach. Note that, although shown in the picture, measurements hitting the borders of the black triangular regions at the bottom are not considered for the likelihood calculation.

5) Penalty Terms: Penalty terms are necessary in order to constrain the particle movements into plausible search space regions. The first imposed restriction refers to the lane width. It is important to allow the variability of this parameter to cope with variations on road construction or with measurement errors (e.g. caused by inaccuracies in the calibration). On the other hand, we are not interested in the whole range of parameters that can be represented by our model. In other words, we want to avoid very large, or very small lane width variants. To avoid these situations we propose the use of a sigmoid function to impose a parameter boundary into our likelihood function. We define the width penalty $\gamma^i_w$ as a function of particle $p^i$ as:

$$
\gamma^i_w = \frac{1}{1 + e^{(w^i - w_{nom})^2 - \delta^2_{cut}}},
$$

where $w_{nom}$ is the nominal lane width (about 4m), $w^i$ is the width component of the given particle, $\delta_{cut}$ is the allowed variation. This function will force 50% penalty at $w_{nom} \pm \delta_{cut}$. We select 2m as the cut point.

A second penalty term refers to the lane parallelism. It represents a strong cue while filtering non-plausible hypotheses, once measurements from both boundaries contribute to reduce the uncertainty. However, in several road situations (e.g. split and merge conditions, or due to calibration variability) the parallel marker assumption is broken, which
requires relaxing this constraint. A system capable of handling both situations would be ideal. In this work we tackle this problem in two steps. We first relax the parallelism constraint by defining different curvature parameters for each lane boundary, as discussed in Section II-A. Secondly, we reintroduce a softer constraint through a penalty term, which favors parallel lanes. The effect of this approach is that the system only allows non-parallel lane models if consistent non-parallel boundaries are present in the measurements. The lane parallelism penalty term is defined as

\[ \gamma_c^i = \exp\left(-\frac{(c^i_l-c^i_r)^2}{2\lambda_c^2}\right), \tag{8} \]

where \(c^i_l\) and \(c^i_r\) are respectively the left and right curvature terms associated to the given particle. The standard deviation parameter \(\lambda_c\) can be interpreted as a degree of confidence of the presence of parallel lanes. By increasing \(\lambda_c\), the parallel constraint is relaxed, while decreasing \(\lambda_c\) forces only parallel lanes to be considered.

6) Likelihood: The final likelihood model is defined by combining equations (6) to (8).

\[ p(z|X^t) \approx \left( \max_{j \in M} \eta^{r,j}_{t,i} + \max_{j \in M} \eta^{l,j}_{t,i} \right) \gamma_{w,t} \gamma_c^i, \tag{9} \]

where \(M\) is the total number of template classes. Left and right boundaries are treated separately. The cue with the maximum weight response is considered for each boundary.

C. PDF Evaluation and Resampling

The estimated PDF over the lane parameters is often non-Gaussian, or even more importantly, is often multi-modal. A typical case is a lane splitting situation, where more than one peak should be present, indicating multiple plausible lanes. In the presence of mismatching patterns (e.g. tire tracks), the effect is similar, and can also be supported over time. We illustrate this property in Figure 4 for a lane splitting situation, where the curvature component of the model forms two distinct modes in the resulting PDF. An evaluation process that discards the multi-modality property of this parameter space, tends to deliver incorrect outputs. A typical approach falling into this mistake is to simply select the output lane as a weighted average of the entire particle set. This makes the estimation lie between the real modes, which is an erroneous result.

A good solution needs to account for the individual PDF peaks separately and postpone the decision. To segment different PDF modes, in our work we apply a slightly modified version of the Fast Adaptive Mean-Shift method [8]. This algorithm adaptively computes the window size applied on the averaging operations. We propose the use of the particle weight as part of the weight term applied to the computation of the mean-shift vector. Samples are normalized before the mean-shift process to account for the inhomogeneous scaling between different dimensions. The algorithm provides as result an association of each particle to a given convergence point, located at the respective mode. The set of derived peaks are therefore better estimations of the final lane parameters. Furthermore, the resulting modes are pruned in respect to their absolute weights (a weak threshold) in order to exclude candidates with non-representative support.

D. Resampling and Update

Particle resampling follows the original SIR [9] implementation, with a single remark that this process is executed only when at least one representative boundary is detected. This avoids resampling while in pure noise situations.

In the update step, ideally particles should be moved to account for the vehicle motion between two steps, considering sensor measurement errors. In our case we treat motion as a component of the measurement noise. Particles are then updated applying a simple additive noise model, based on a Gaussian distribution in each component dimension.

III. Evaluation

The proposed system was evaluated in a 11-minutes sequence recorded at 15Hz with a Guppy F-036C (Allied Vision Technologies) camera, equipped with an Aptina MV022 automotive sensor. The scenes contain a typical driving scenario in highway roads. This includes different situations in terms of measurement and layout: straight roads, curves, lane splitting and merging, exits, different illumination conditions, including saturated frames, occlusions, shadows and mismatching patterns. From the entire sequence a set of interesting (i.e. challenging for a lane detection task) segments were extracted and labeled. For straight roads, labels were applied every second, while for more complex scenarios labels were introduced every half second. A total of 117 frames were manually labeled.

A. Error Metric

Two different metrics were selected for the evaluation. The first captures if the markers are detected, missed or if false positives occur. This is accomplished by computing the distance in image space between the detected modes and the ground truth data. A boundary is declared detected if the average error between the estimate and the ground truth label is below 1.0m. Labels with no association are declared as misses, while detected boundaries not associated with any label are considered as false positives.
TABLE I
QUANTITATIVE RESULTS

<table>
<thead>
<tr>
<th>Scenario</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Boundaries</td>
<td>60</td>
<td>16</td>
<td>29</td>
<td>10</td>
<td>20</td>
<td>79</td>
</tr>
<tr>
<td>Positive Matches</td>
<td>59</td>
<td>15</td>
<td>29</td>
<td>10</td>
<td>16</td>
<td>57</td>
</tr>
<tr>
<td>Misses</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>False Positives</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Average RMSE</td>
<td>0.193</td>
<td>0.197</td>
<td>0.236</td>
<td>0.251</td>
<td>0.313</td>
<td>0.382</td>
</tr>
</tbody>
</table>

The second metric is the effective error between detected modes and ground truth labels. For every meter in road space, the RMSE between a detected boundary and its corresponding label is computed. Single measurements are averaged representing the boundary error (in meters).

B. Results and Discussion

We separate our results in six sub-sequences:

1) Straight road with variation of the pitch angle;
2) Lane change on straight road with high pitch variation;
3) Lane change with low road curvature to the left;
4) Under bridge driving (shadow and glare);
5) Under bridge driving with occlusion by near vehicles;
6) Exit scenario, with wide and splitting lane;

Table I shows quantitative results of the proposed methods. Evaluation of sequences 1 to 3 show that the proposed approach responds well to the most common scenarios found in a typical highway environment. One single boundary was missed in the first sequence, resulting from the degraded quality of the middle dashed marker (Fig. 5(a)). The method showed to robustly handle rather strong mismatching patterns originated from tire rails. In the second and third sequences the few false positives conditions occurred during the lane changing maneuver. This is mainly a consequence of the natural latency introduced by the movement of particles among different modes. The missing boundary (Fig. 5(d)) results from a weak support of the respective mode caused by three factors. First, the absence of motion compensation in the Particle Filter update step, may lead the target mode to fall onto regions with sparse particle coverage. Second, in the current setup, only black-white-black patches contribute to the likelihood estimate, imposing a natural penalty to dashed markers. And third, the presence of mismatching patterns (noise) increases the uncertainty, therefore influencing the mode estimation. After lane changing, the hypothesis representing the previous lane also looses support, due to the fact that fewer patches contribute to the estimate. Figures 5(f) and 5(e) shows the estimations respectively before and after the lane change maneuver. Also important to remark is the robustness of the system face to high variation of the pitch angle present on the sequences. Partially relaxing the parallelism constraint contributes significantly to this result.

The analysis of the fourth sequence reflects the robustness of the HoG descriptor to illumination variance (see Figures 5(g) and 5(h)). The fifth scenario adds to the illumination challenges, partial occlusion of markers, vehicle shadows and poor contrast as can be seen in Figures 5(i) and 5(j). These factors strongly degenerate the likelihood estimate, leading again to a poor support of the target model during some frames. The last scenario (Figures 5(k) and 5(l)) represents an exit with two split conditions, where the right lane becomes wide enough to accommodate two new lanes. Although the system is able to recover in the split condition, particles do not migrate fast enough to form new modes immediately, so as to follow both lanes.

RMSE values are close related to the uncertainty present in the respective scenes. Higher confidence levels lead to a bigger concentration of particles around the true model, increasing the precision of the estimation. Exactly the opposite occurs when uncertainty grows. Another factor contributing to the increasing error in more complex scenarios is the amount of mismatching patterns distracting the filter. Particles covering these regions can be wrongly included in the segmentation and the computation of the real mode, propagating the errors to the final estimation.

IV. CONCLUSIONS AND FUTURE WORK

In this work we introduce the concept of feature templates as the basis for a new likelihood model in particle filter based lane detection. Features extracted at selected grid locations are compared to descriptors pre-stored in a dataset. Pairs of extracted/template descriptors are selected from particle hypotheses, and the resulting distance is applied for the likelihood estimation. Such an approach allows a scalable evaluation of multiple patterns against the extracted feature vector. To derive the models from the resulting PDF, a Mean-Shift segmentation is applied. This approach better captures the natural multi-modal characteristic of the lane detection task. Through the evaluation we have shown that, even using simple synthetic templates, the method is able to handle with efficiency standard highway scenarios.

These promising results lead to a new set of possibilities regarding lane and road boundaries discrimination, ranging from simple evaluations of different feature descriptors and metrics up to the inclusion of state-of-the-art machine learning approaches. The synthetic templates can intuitively be replaced by manually annotated captured data. This allows the inclusion of more complex patterns (e.g. curbs, dot bots, guard rails) directly into the Particle Filter estimation step. In our future work we will target the inclusion of Support Vector Classification and Regression methods in the likelihood model. We will also target the improvement of the segmentation step, in order to better retrieve the estimated models, especially in terms of normalization of particle dimension ranges and on the determination of the Mean-Shift bandwidth size.

REFERENCES

Fig. 5. Resulting lane detection for the six different scenarios. From top left to bottom right, two examples for each scenario. Left and right components of the model are projected onto the Bird’s Eye View image, in respectively red and magenta. Figure to be visualized in color.


