COMBINED DYNAMIC TRACKING AND RECOGNITION OF CURVES WITH APPLICATION TO ROAD DETECTION

Jean-Philippe Tarel

LIVIC (INRETS-LCPC)
13, Route de la Minière
78000 Versailles, France

Frédéric Guichard[*]

Poseidon / Vision IQ
3, rue Nationale
92100 Boulogne, France

ABSTRACT

We present an algorithm that extracts the *largest* shape within a specific class, starting from a set of image edgels. The algorithm inherits the Best-First Segmentation approach [1]. However, instead of being applicable only to shapes defined within a given class of curves, we have extended our approach to tackle more general - and complex - shapes. For example, we can now process shapes obtained from sets defined over different kinds of curves and related to one another by estimated parameters. Therefore, we go from a segmentation problem to a recognition problem. In order to reduce the complexity of the searching algorithm, we work with a linearly parameterized class of shapes. This allows us, first, to use a recursive Least-Squares fitting, second, to cast the problem as the search of a largest edgel subset in a directed acyclic graph, and, third, to easily introduce a priori information on the location of the edgels of the searched subset. This leads us to propose a unified approach where recognition and tracking are combined. Experiments on recognizing and tracking both left and right road boundaries demonstrate that real-time processing is achievable.

1. INTRODUCTION

A robust detection and tracking - via on-board camera - of road lane markings and boundaries is of major importance for automatic vehicle guidance. For lateral vehicle guidance, road boundaries detection must at least provide at a good rate estimates of the relative orientation and of the lateral position of the vehicle with respect to the road. Automatic vehicle guidance has been a subject of investigations from many years [2, 3]. Contrary to these techniques and many classical approaches, we merge the recognition and the tracking processes into one single process. In addition, in order to cope with occlusion by vehicles, signs, light spots or shadows, and low image contrast, we do not assume particular markings or road lightning conditions. Lane recognition starting from a large set of edgels is based only on geometrical features.

[*]The second author performed the work while at LIVIC

This paper presents an approach based upon the curve segmentation algorithm first introduced in [1]. This algorithm proves especially attractive because it can handle curves even in situations where there are gaps in the data, this in an automatic fashion. In the context of this algorithm, shapes in 2D images are described by their boundaries, which are then represented by linearly parameterized curves. In [1], the N longest curves which individually fit into a chosen class of curves are obtained. Due to the geometry of the road, the probability to obtain the lane boundaries from the N longest curves proves to be high. However, no consistency between extracted curves is imposed neither in time nor in space.

We therefore extend this method: (i) by modeling lane boundaries as the image of two parallel curves on the road, so that left and right boundaries contribute to that same model; (ii) by imposing time consistency between the extracted models, so that the left and right boundaries are simultaneously extracted. This provides reliability in the detection - the probability to find two parallel curves that do not correspond to the lane boundaries is smaller than for two independent curves -, better measurement since the two curves are embedded into the same model, and a reduction of the computation time thanks to the tracking. An alternative way to think about our algorithm, is that it consists in a model based tracking combined with extracting the largest data set fitting the model.

2. RECOGNITION

2.1. The Model

We assume that the road is planar and that its boundaries may be approximated by a polynomial of degree d: $y^* = \sum_{i=0}^{d} b_i \; {x^*}^i$. The transformation between the road plane (x^*, y^*) and the image plane (x, y) is: $x = l_x \frac{1}{x^*}$ and $y = l_y \frac{y^*}{x^*}$, where l_x and l_y are only functions of the camera calibration parameters. We set the origin of the coordinate system to a point on the *sky line* - the position of this line can be computed from the camera calibration. Then, the left road boundary is projected in the image as curve s_{left}:

$$y = \sum_{i=0}^{d} a_i x^{1-i} \tag{1}$$

The right road boundary is assumed to have the same shape as the left one, and being translated by the width of the lane. Therefore, its projection in the image is the curve s_{right}:

$$y = \sum_{i=0}^{d} a_i x^{1-i} + wx \tag{2}$$

where w is the unknown width of the road lane. In a typical image, the road boundaries appear as pieces of s_{left} and s_{right}.

We define the extended road boundaries model on the basis of the above pairing of curves. The extended parameters of such a shape are $A = (a_i, w)$, $0 \leq i \leq d$. Defining $F_{left}(x) = (x, 1, x^{-1}, \ldots, x^{1-d}, 0)$ and $F_{right}(x) = (x, 1, x^{-1}, \ldots, x^{1-d}, x)$, equations (1) and (2) above become $y = A^t F_{left}(x)$ and $y = A^t F_{right}(x)$, respectively.

2.2. Maximum Length Criterion

We formulate the "lane boundary detection problem", as a search for the longest boundaries that belong to the shape model introduced above. Since the lane boundary model is a linearly parameterized class of shapes, it can be detected by extending the searching algorithm we proposed in [1]. In this earlier work, the highest energy subsets of edgels that fit to an underlying curve are recursively detected, see Fig. 1(c). Following the Mumford and Shah [4] approach, the energy was defined as a tradeoff between the fitting error and a function e^{over} of the length of the edgel subset $\{p_1, \ldots, p_m\}$:

$$\mathcal{E}(p_1, \ldots, p_m) = \frac{e^{fit}(p_1, \ldots, p_m)}{\lambda} - e^{over}(p_1, \ldots, p_m) \tag{3}$$

A difficult problem in obtaining such an energy is the choice of the parameter λ. Indeed, the measure e^{over} is balanced with the fitting error and therefore its convexity is related to the fitting error convexity, as well as λ, and it cannot be physically interpreted without an explicit definition of e^{over} and e^{fit}.

We have found experimentally, that it is better to accept an edgel subset as a valid approximation of an underlying parametrized shape, if its e^{fit} is lower than a threshold λ. The value of λ is then directly related to the observed accuracy of the shape drawing in the image. The first term in the energy (3) is changed accordingly by $g(\frac{e^{fit}(p_1, \ldots, p_m)}{\lambda})$, where g is a step function defined as $g(x) = 0$ when $x \leq 1$ and $g(x) = +\infty$ anywhere else. The selection of a step function g permits us to bypass the delicate choice of a balance between the two terms in (3). Here g forces the data to be approximated within a range from 0 to λ. Hence, regardless of the choice of e^{over}, the problem becomes simply to find the largest edgel subset which approximates a shape in a given class. Roughly speaking, instead of searching for the best tradeoff between e^{fit} and e^{over}, we are now searching for the largest subset, *i.e.*, the one that maximizes e^{over}, among those which satisfy a minimal quality fitting criterion, *i.e.*, $e^{fit} \leq \lambda$. We call this formulation the *Maximum Length Criterion* under a fitting constraint, or MLC in the following.

Choosing simply, for MLC, the sum length of edgels in the subset, the problem can be equivalently cast as the minimization of:

$$\mathcal{E}(p_1, \ldots, p_m) = g(\frac{e^{fit}(p_1, \ldots, p_m)}{\lambda}) - \sum_{i=1}^{m} l_i \tag{4}$$

where l_i is the length of edgel p_i. From (4) and following the Best-First Segmentation approach [1], one can naturally make use of the MLC for partitioning all edges of a given image in subsets, where each subset approximates a shape in a given class. Indeed, the Best-First Segmentation approach is always applicable, since it consists in finding the largest edgel subset first, followed by a removal of this subset from the edgel set. Then one may iterate the search for the next largest edgel subsets, and so on, until there are no edgels left.

In the context of edge segmentation, on the contrary to approaches based on the energy (3), the MLC yields a problem controlled simply by the λ parameter, which has a physical meaning - the maximal average distance approximation. Furthermore, the MLC can be chosen as need fits, since it will not be balanced with the fitting error.

2.3. Recognition Algorithm

The recognition process in an image consists in (see [1] for details):

- **Edgel Detector**: For minimizing as much as possible the use of contrast-based selections, an edgel is defined as a straight line segment embedded in level lines of the given image. Since the thresholding of the gradient magnitude is reduced to a single gray level difference, the edgels are numerous, see Fig. 1(b), but contrary to other approaches no low contrast boundaries are missed.
- **Graph Building**: Edgels are ordered by their coordinates and organized as nodes in an acyclic directed graph, i.e every edgel is linked to all the follow-

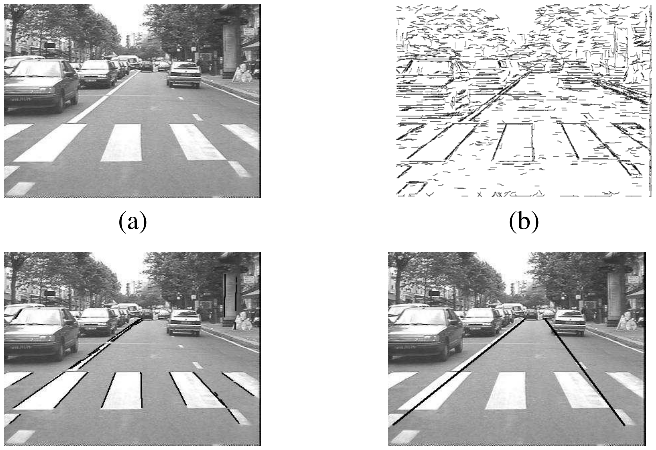

Fig. 1. (a) the original image and (b) the result of the straight line segment detector. From (b), (c) the segmentation in 15 best curves and (d) the longest shape modeling both left and right road boundaries with extended parameters using tracking.

ing *consistent* edgels. The consistency depends upon proximity and alignment of both edgels.

- **Graph Searching**: The distance between an edgel and both curves s_{left} and s_{right}, is the minimal distance of this edgel to each of these curves. Then any edgel is associated to the curve with the lowest distance for recursive update of A parameters. By choosing such a distance between an underlying lane boundary and an edgel, we allow Least-Squares minimization with respect to the shape parameters. Thus, we directly apply recursive Least-Squares theory, i.e the main property upon which Kalman filtering is based. Recursive fitting is used with great advantage during the graph searching to perform fast pruning. The idea is to keep not only the best assumption arriving at the current node, such as in dynamic programming, but an ordered list of the b best assumptions. An assumption is here an edgel subset represented by its underlying shape parameters.

This algorithm can be applied using the shape parameters $A = (a_i)$ or the extended ones $A = (a_i, w)$. Fig. 1 illustrates the process: the original image (a) is converted into a list of edgels (b). In (c), we display the 15 longest curves estimated independently. We see that even if the left side of the lane is correctly extracted, the right side is not long enough with respect to the pedestrian crossing markings and therefore is not extracted. Whereas by using in (d) the extended shape parameters, i.e the lane boundaries are a pair of curves, the lane boundaries are extracted as the first longest pair of curves having time consistency, as described next.

3. DYNAMIC TRACKING

3.1. Initializing by Bias Towards a Simpler Shape

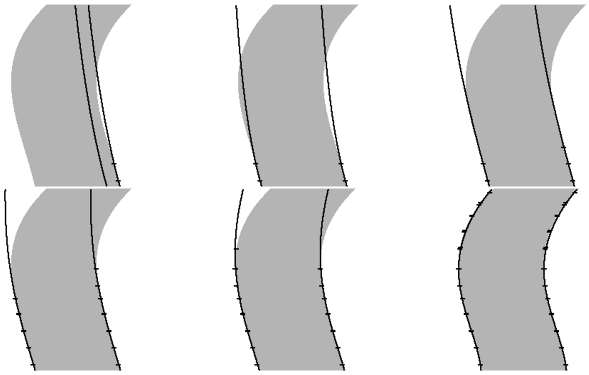

Fig. 2. Recursive fitting with an increasing number of edgels (1, 2, 3, 9, 11, and 20 respectively) by parallel pair of curves of degree 5. There is 7 parameters for this kind of shape. When the number of edgel is too small to constrain enough the shape parameters, the result is biased towards parallel lines.

In the first steps of the graph searching, under-constrained fitting of parametrized curves is performed on small sets of edgels. When an edgel set contains less edgels than the degree of freedom of the fit, instabilities occur. But the graph building requires implicitly the initialization of fitting parameters A_0 and covariance matrix K_0 for each single edgel. Following the Kalman's theory, by choosing these initial values, we have the chance to control the a priori information on the possible shapes and locations of the searched curves.

Mainly, the choice of a specific A_0 and K_0 depends on the application context. For instance, for the recognition of road boundaries shape in the first image of a video sequence, we set $A_0 = (0, \dots, 0, u)$, where u is the width of a standard lane, and K_0 equals a diagonal matrix with diagonal $(0, v, \frac{v}{2}, \dots, \frac{v}{d}, v)$. The value of v is chosen to be large enough to allow the shape to be able to deform and thus fit the provided data. v is the a priori variance of the lane width w, the last shape parameter. There is zero for the a priori variances of a_0, since we want the fitting to be invariant to a y translation of the road boundaries. We chose decreasing a priori variances for a_i parameters, because we want to a priori bias the resulting shape towards a lower degree shape such as a pair of converging lines rather than a pair of higher degree curves.

In Fig 2, the intermediate fits illustrates how the fitting, and thus the recognition, is biased towards parallel lower degree polynomial. When the data does not contain enough information for accurate estimates, the algorithm fits the data set by two curves close to two parallel lines. Notice the robustness of the fitting to the bad initialization of the lane width w.

Lower are the variances in K_0, i.e better is our confidence in the a priori information, more the fitting error e_{fit} discriminates from edgels close to A_0 to edgels far away. As a consequence, better is the confidence, lower can be the number b of best assumptions we need to keep at each node during the graph searching, and thus faster is the algorithm. This leads to efficient dynamic tracking of shapes within rather complicated class of shapes.

3.2. Time consistency

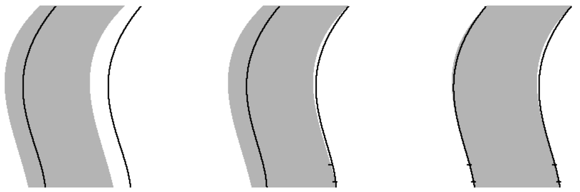

Fig. 3. Recursive fitting with an increasing number of points by parallel pair of curves of degree 5. The used a priori information is a translated version of the searched curves (to be compared to Fig. 2).

Let t being the time index of the current image, we assume to know parameters A_{t-1} and covariance matrix K_{t-1} of the shape in the previous image. From A_{t-1} and K_{t-1}, we want the algorithm to search a shape at time t in a location close to and with a shape similar to the one at time $t-1$. Fig 3 shows how using previously detected curves the algorithm converges close to the optimal solution when only two edgels are processed.

When a dynamic model of the vehicle is known, the initialization of A_0 and K_0 can be obtained using predictions for A_t and K_t. Contrary, when this model is unknown, a *Principal component analysis* can be performed on the parameters obtained from pre-processed videos of road lane-markings. This allows a learning on how the shape is usually moving and deforming. This is used then to set K_0 to a rescaled version of K_{t-1}, where scaling is applied only in the directions of main variation of the curves parameters.

This initialization introduces not only a bias toward the previous extracted shapes, but will also add a priori information on the area where to search for edgels. We are in fact implicitly focusing the search in areas where interesting edgels has a high probability to be found, and therefore we speed up the searching. This allows the process to run in a reasonable time on rather complicated images and permits efficient simultaneous recognition and tracking.

Fig. 4 shows tracking results of road boundaries in images with light spots or with holes in the lane-markings. In Fig. 1 computation time on a Pentium 200Mhz, 32Mo are: (b) $0.04s$ for the edgel set, (c) $1s$ for the best 15 curves without tracking, (d) $0.3s$ for best pair left and right boundaries with tracking.

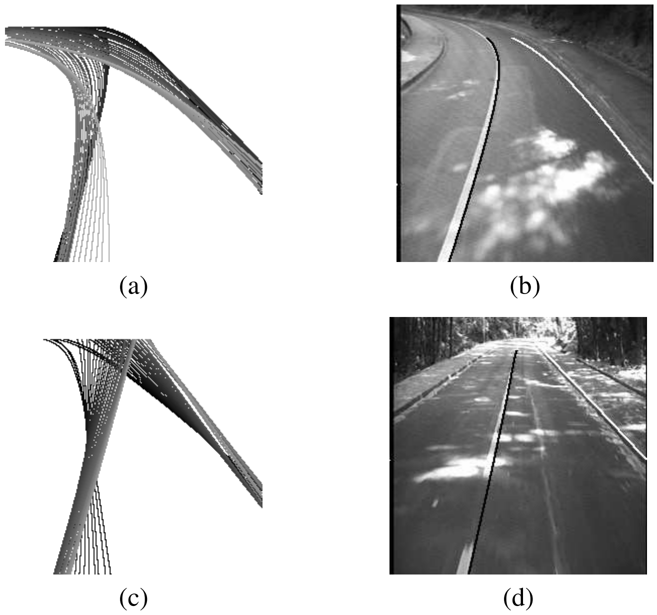

Fig. 4. (a) (c): the largest pair of left and right road boundaries is tracked in 40 images. (b) (d): one of the processed image where light spots and shadows may induce recognition difficulties.

4. CONCLUSION

We were able to merge the tracking with a recognition algorithm which ensures that the largest data set is matched with the parametric model of the target. On one hand, the use of extended parameters increases the searching complexity, but, on the other hand, the use of *a priori* information from the tracking allows us to retrieve the lane boundaries as the firstly extracted pair of curves. The proposed approach may provide a useful framework in other contexts and we are currently working on extending it even more for tracking and recognizing closed curves in images.

5. REFERENCES

[1] F. Guichard and J.-P. Tarel, "Curve extraction combining perceptual grouping and a kalman like fitting," in *ICCV*, Kerkyra, Greece, 1999.

[2] R. Chapuis, A. Potelle, J.L. Brame, and F. Chausse, "Real-time vehicle trajectory supervision on the highway," *IJRR*, vol. 14, no. 6, pp. 531–542, 1995.

[3] C. Thorpe, M. Herbert, T. Kanade, and S. Shafer, "Vision an navigation for the canegie-mellon navlab," *T-PAMI*, vol. 10, no. 3, 1988.

[4] D. Mumford and J. Shah, "Boundary detection by minimizing functionals," in *CVPR*, 1985, pp. 22–26.