

3D Data Driven Prediction for Active Contour Models with Application to Car Tracking

Joachim Denzler and Heinrich Niemann
Lehrstuhl für Mustererkennung (Informatik 5)
Universität Erlangen-Nürnberg

Abstract

In this paper we apply a new data driven 3D prediction step for active contour models to car tracking on highways. The so called 3D bounding volume (BV) is a coarse 3D representation of a moving object, for which the 2D contour in the image plane has been extracted and tracked by active contours. By calculating the BV's shape and location in 3D an estimation of the object's motion is possible. Thus, in contrast to pure 2D tracking of the object's contour by active contour models knowledge about 3D motion is available. This is necessary, if changes in the object's contour — for example, due to rotation — needs to be predicted.

We present experiments in the area of car tracking, which show that tracking the cars by active contour models can be improved by the proposed 3D prediction step. In addition, relative statements about the direction of the motion and the velocity of the cars are possible.

1 Introduction

In the past years active contour models have been successfully applied to object tracking. Despite the fact that for object tracking a prediction step is an essential part, only few work is known which introduces a 2D prediction step into the framework of active contours. For example, [2] computes a 2D prediction based on the normal flow measured at the snake elements in the image. [12] proposes a Kalman-snake which is capable for tracking 2D contours.

There is one main reason for the lack of a 3D prediction: For tracking moving contours, a prediction of an object's contour is only possible if 3D knowledge about the object itself is available. Due to the fact that active contour models are applied to data-driven tracking no model knowledge is normally available.

In some application a coarse idea about the objects is available, without having an explicit representation. For example, so called generic car models have been used to track cars in traffic scenes

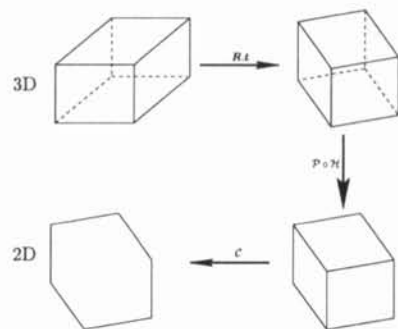


Fig. 1: BV for 2D contour prediction: The mappings $\mathcal{P}$, $\mathcal{H}$, and $\mathcal{C}$ are formally described in Sect. 2.1

[8, 10]. the explicit parameters of the car model are estimated during the tracking itself. Another example is the generic model of humans for pedestrian tracking [11]. This principle is transferred in this paper to active contour models to introduce a data driven 3D prediction step. For data driven tracking where no a-priori models of the object are available one has to look for a description of the object which enables to predict the 2D contour of the object by estimating the 3D position and the coarse shape of the object itself. The bounding volume of an object, which is a well known term in computer graphics, is the smallest volume which completely contains the object. These bounding volumes can be applied to 2D contour prediction. The idea is the following (see also Fig. 1): Initially extract the contour of the moving object by the snake's energy minimization, then estimate the parameters of the BV (i.e. the location in 3D and its shape), such that the projected contour of the BV best matches the extracted active contour. Finally, use the computed location and shape in 3D to update 3D knowledge about the motion and the shape of the object. For the next image the contour of the BV is projected into the 2D image plane to initialize the active contour.

In Sect. 2 the approach of 3D bounding volumes (BV) will be introduced. We also present the mo-

tion model and the estimation algorithm, which have been applied in the experimental part of this paper (Sect. 3). There, experiments in the area of car tracking on highways show that object tracking by active contours can be improved and even relative statements about the direction of the motion and the velocity of the cars are possible. The paper closes with a discussion of the results and an outlook to future work (Sect. 4).

2 Theoretical Background

2.1 3D Bounding Volume

Due to lack of space, we only shortly summarize the idea of the BV. A more detailed description can be found in [4]. Let $M(\mathbf{a})$ be the set of 3D points of a BV, parameterized by a vector $\mathbf{a}$:

$$M(\mathbf{a}) = \left\{ ({}^w x_i(\mathbf{a}), {}^w y_i(\mathbf{a}), {}^w z_i(\mathbf{a}))^T \mid i = 1, \dots, n \right\}$$

The upper left w denotes that the coordinates ${}^w x_i$, ${}^w y_i$ and ${}^w z_i$ of the point i refer to the 3D world. These points may be corners, edge points or in general surface points of the BV. For example, for a rectangular solid, shown in Fig. 1, a parameter vector $\mathbf{a}$ might be $\mathbf{a} = (l, w, h)^T$, with l, w and h being the length of the edges of the rectangular solid. In general no restrictions for the object's shape are made. The rotation $\mathbf{R}$ and the translation $\mathbf{t}$ map the points of $M(\mathbf{a})$ to the set ${}^{\mathbf{R}, \mathbf{t}} M(\mathbf{a})$, which contains the rotated and translated 3D points of the BV. Now, a visibility test must be performed. In the literature of computer graphics several algorithms can be found (z -buffer, scan-line, raytracing [6]). We define a hiding operator $\mathcal{H}$, which maps the set ${}^{\mathbf{R}, \mathbf{t}} M(\mathbf{a})$ of 3D points into the set ${}^{\mathbf{R}, \mathbf{t}} M'(\mathbf{a})$ of visible 3D points. Now, the set ${}^{\mathbf{R}, \mathbf{t}} M'(\mathbf{a}) \subset \mathbb{R}^3$ will be projected onto the image plane by perspective projection $\mathcal{P}$. The result is the set ${}^{\mathbf{R}, \mathbf{t}} M_{\mathcal{P}}(\mathbf{a})$ which is equal to the 2D image of the BV's points. Finally, an operator $\mathcal{C}$ will compute the visible 2D contour of the BV, which leads to a set of points ${}^{\mathbf{R}, \mathbf{t}} C_{\mathcal{P}}(\mathbf{a})$ in $\mathbb{R}^2$. These points need to be transformed to a sequence of points $\langle c_i \rangle_{1 \leq i \leq m}$, with $c_i \in {}^{\mathbf{R}, \mathbf{t}} C_{\mathcal{P}}(\mathbf{a})$, ordered counterclockwise to form a representation of this contour.

In Fig.1 all steps of this approach are summarized. The mappings $\mathcal{H}$ and $\mathcal{C}$ are time critical for real-time experiments. By taking as BV the special class of convex polyhedra these both mappings can be done by projecting the corners of the convex polyhedra into the image plane and calculating the convex hull of these points. This computation is obviously less time consuming and can be applied to real-time problems.

For two contours $\langle c_i \rangle_{1 \leq i \leq m}$ and $\langle c'_j \rangle_{1 \leq j \leq n}$ a distance function $\text{dist}(\langle c_i \rangle_{1 \leq i \leq m}, \langle c'_j \rangle_{1 \leq j \leq n})$, for example

$$\text{dist}(\langle c_i \rangle_{1 \leq i \leq m}, \langle c'_j \rangle_{1 \leq j \leq n}) =$$

type of vehicle	length	width	height
car	4.0	2.0	1.5
van	5.0	2.5	2.5
truck	12.0	3.0	3.0

Tab. 1: Parameters of the BV for different cars.

$$= \sum_{i=1}^m \min_{c'_j} \{ |c_i - c'_j| \} + \sum_{j=1}^n \min_{c_i} \{ |c_i - c'_j| \} \quad (1)$$

is defined. This function measures the correspondence of two 2D contours. Now, for a given active contour $\langle c'_j \rangle_{1 \leq j \leq n}$ and a parameter description of a BV, the parameters $\mathbf{R}, \mathbf{t}$ and $\mathbf{a}$ can be computed by

$$(\mathbf{R}, \mathbf{t}, \mathbf{a})^T = \underset{\mathbf{R}, \mathbf{t}, \mathbf{a}}{\text{argmin}} \text{dist}(\langle c_i \rangle_{1 \leq i \leq m}, \langle c'_j \rangle_{1 \leq j \leq n}) \quad (2)$$

where $c_i \in {}^{\mathbf{R}, \mathbf{t}} C_{\mathcal{P}}(\mathbf{a})$. The minimization in (2) results in the parameters $\mathbf{R}, \mathbf{t}$ and $\mathbf{a}$ of that BV, the contour of which best matches — in the sense of equation (1) — the active contour. Of course, ambiguities especially for the parameter $\mathbf{R}$ may occur (the Necker illusion); in that case, local minima may be reached. The experiments will show, that these local minima are no problems for the prediction of the contour. To calculate the parameters $\mathbf{R}, \mathbf{t}$ and $\mathbf{a}$ we use stochastic optimization techniques described in [5].

After this step we have a 3D estimate of the moving object's BV. The only knowledge which is needed for this step is a parametric representation of the BV, which has to be chosen in advance. In our experiments (see Sect. 3) we have taken a rectangular solid.

2.2 Motion Model and Prediction

With the algorithm presented in the previous section we can calculate for each 2D active contour the shape and location of a BV, which 2D contour best matches the active contour. Now, in the case of image sequence processing we get for each image $f(x, y, t)$ the parameters $\mathbf{R}(t), \mathbf{t}(t)$ and $\mathbf{a}(t)$. Thus, an estimation of the shape parameters and the motion of the BV in 3D is possible. Usual approaches can be found in estimation theory [1].

Despite the fact, that the parameter vector $\mathbf{a}$ of the BV can also be estimated as described in the previous section, we use for the experiments only three different parameter vectors $\mathbf{a}$. This reduces the complexity of the search space. The parameter vectors correspond to three different types of vehicles (car, van, and truck) and have been determined heuristically and fixed in advance. The relative parameter values can be found in Tab. 1. It is worth noting, that these values are only coarse estimations.

For the motion model we apply the discrete-time model of a constant-velocity target [1]. The state of the target (position, velocity) is estimated by a Kalman-Filter.

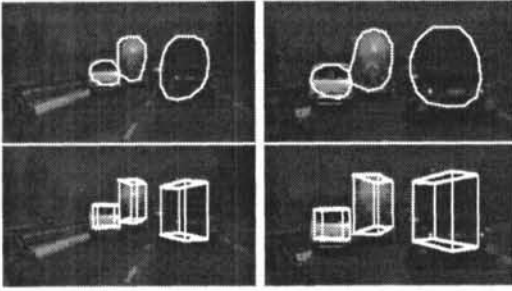


Fig. 2: Results for tracking cars on a highway: the first and the last image of a sequence of 124 images are shown. First row: the extracted active contours. Second row: the estimated BV.

3 Experiments and Results

3.1 Experimental Environment

We have tested our proposed method on highway image sequences (one example is shown in Fig. 2). This data set contains 10 sequences, each with a length of approximately 100 - 200 images. For the first image, each active contour is initialized interactively on the corresponding moving vehicle. This is due to the fact that we have no knowledge about the movement of the camera and are thus not able to estimate independent motion in the image. An automatic initialization in the case of known camera motion has already been proposed in [9].

Then, tracking is done with active contours without any prediction step. We use an active contour model, which is based on the original approach of [7] and which has been modified to fulfill real-time constraints [3].

The image sequences, which have been used in this paper, are very difficult to process with active contour models. The reason for this is that there are background edges near the object (other vehicles), weak object contours (very low contrast), and large displacement of the vehicle in the image plane (especially for vehicles approaching the observer). Thus, normally the active contour loses the moving vehicle after some images.

Once the active contour has lost the object, the second experiment starts. As long as the estimation error of the Kalman-Filter is above a certain threshold, tracking is done without the prediction, i.e. initialization of the active contour. After the Kalman-Filter error is below the threshold, the prediction step by the BV is activated, for which the location in 3D has been already estimated and updated during the previous images. Then, for each new image the 3D location of the BV is predicted and its 2D contour is projected into the image plane. This 2D contour is used to initialize the active contour, which extracts the object contour by the normal energy minimization.

3.2 Results

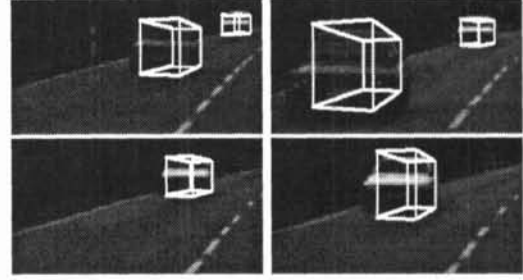


Fig. 3: Tracking a car approaching the observer by the BV. Even the pose estimation is correct.

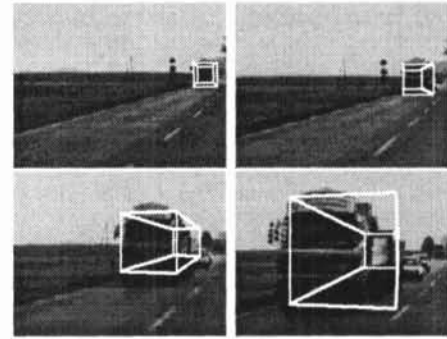


Fig. 4: Tracking a truck approaching the observer.

In our experiments a total number of 13 vehicles have been tracked. The average number of images, in which a vehicle has been visible, is 98 images. Without any prediction only one sequence has been completely tracked without an error. With the proposed 3D prediction step, we were able to correctly track the vehicle over the whole sequence in 6 of the 13 sequences. The average number of images, in which a vehicle could be tracked, was 28 images without prediction and 46 images with the prediction step. For one sequence neither with nor without prediction step the vehicle could be tracked. The reason is, that there is a very low contrast in the image and the distance to the vehicle is large, which results in a very small object contour. It is well known that for such kind of image data active contour models are not suited.

In the following we will illustrate the advantages of the algorithm. As one can see in Fig. 2, the BV does not correctly model the real 3D position of the vehicles. Nevertheless, the computed 2D contour of the BV, which is taken as initialization of the active contour for each new image, is precise enough to track the object correctly.

In Fig. 3 and Fig. 4 two example sequences are shown, for which the tracking without 3D prediction

fails. Even if the large displacement of the contour in the image plane could be estimated, the simultaneous growing of the contour cannot be predicted without a 3D model. As one can see, with the BV prediction step, the vehicles can be tracked correctly.

In Fig. 5 the estimated relative distances for the three vehicles (the van, the tanker, and the truck) are shown. No absolute 3D position can be computed, because no exact model for the vehicles and no calibrated cameras are available. But as one can see, the relative change in the distance corresponds to the movement of the camera towards the three vehicles.

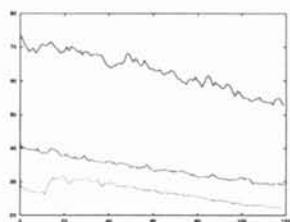


Fig. 5: Estimated relative distance of the three vehicles (see Fig. 2) to the camera over the image sequence by means of the BV. The change in distance corresponds to the movement towards the vehicles.

4 Discussion and Future Work

In this contribution we have shown, that the proposed prediction method for active contour models is well suited to improve the performance of a data driven tracking. Weak object contours, large displacements of the moving object and sudden loss of the object can be handled. The BV itself of course cannot be taken as an exact representation of the moving object, i.e. the BV does not always model the real shape of the object. Nevertheless, the relative motion (in this case the shrinking or growing distance), which is the necessary information for a prediction step, is always modeled exactly. With this information statements about the motion direction and velocity of the object can be made, which is impossible without a 3D estimation.

Up to now, there are several problems. The initialization of the Kalman-Filter parameters is a very difficult task. Thus, if the active contour cannot track the object without prediction sufficiently long, the Kalman-Filter may not be in steady state, and thus predict a wrong motion. In that case, as for each wrong initialization of the active contour, the object gets lost.

A second problem occurs, when the active contour slowly loses the object's contour. Then, the Kalman-Filter will predict an increasing distance of

the object (due to a shrinking contour) or some rotation of the BV, which is not correct. As a result, a wrong contour is predicted and the object cannot be tracked any longer.

In our future work, we will also estimate the shape of the BV during the tracking, instead of using fixed values. Furthermore, some other motion models of the cars will be tested.

References

- [1] Y. Bar-Shalom and T.E. Fortmann. *Tracking and Data Association*. Academic Press, Boston, San Diego, New York, 1988.
- [2] M. Berger. Tracking rigid and non polyhedral objects in an image sequence. In *Scandinavian Conference on Image Analysis*, pages 945–952, Tromsø (Norway), 1993.
- [3] J. Denzler and H. Niemann. Evaluating the performance of active contour models for real-time object tracking. In *Second Asian Conference on Computer Vision*, pages II/341–II/345, Singapore, 1995.
- [4] J. Denzler and H. Niemann. 3d data driven prediction for active contour models based on geometric bounding volumes. *Pattern Recognition Letters*, page to appear, 1996.
- [5] S. M. Ermakov and A. A. Zhiglyavskij. On random search of global extremum. *Probability Theory and Applications*, 28(1):129–136, 1983.
- [6] J.D. Foley, A. van Dam, S.K. Feiner, and R.L. Phillips. *Grundlagen der Computergraphik*. Addison-Wesley, Bonn, Paris, Reading, Massachusetts, 1994.
- [7] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 2(3):321–331, 1988.
- [8] D. Koller, K. Daniilidis, T. Thorhallson, and H. Nagel. Model-based object tracking in traffic scenes. In G. Sandini, editor, *Computer Vision - ECCV 92*, pages 437–452, Berlin, Heidelberg, New York, London, 1992. Lecture Notes in Computer Science.
- [9] D. Murray and A. Basu. Motion tracking with an active camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5):449–459, 1994.
- [10] H. Niemann, W. Kasprzak, and P. Weierich. Integrated motion and geometry based obstacle detection in image sequences of traffic scenes. In *Proceedings 10th Annual Symposium AeroSense'96*, Orlando, FL, 1996.
- [11] J.M. Rehg and T. Kanade. Model-based tracking of self-occluding articulated objects. In *International Conference on Computer Vision*, pages V/612–V/617, Cambridge, MA, 1995.
- [12] D. Terzopoulos and R. Szeliski. Tracking with Kalman snakes. In A. Blake and A. Yuille, editors, *Active Vision*, pages 3–20. MIT Press, Cambridge, Massachusetts, London, England, 1992.