Robust Multiple Car Tracking with Occlusion Reasoning

Dieter Koller, Joseph Weber, and Jitendra Malik

University of California at Berkeley,
EECS Dept., Berkeley, CA 94720, USA
Email: {koller,jweber,malik}@eecs.berkeley.edu

Abstract. We address the problem of occlusion in tracking multiple 3D objects in a known environment. For that purpose we employ a contour tracker based on intensity and motion boundaries. The motion of a contour enclosing the image of a vehicle is assumed to be well describable by an affine motion model with a translation and a change in scale. Contours are represented by closed cubic splines the position and motion of which are estimated along the image sequence. In order to employ linear Kalman Filters we decompose the estimation process in two filters: one for estimating the affine motion parameters and one for estimating the shape of the contours of the vehicles. Occlusion detection is performed by intersecting the depth ordered regions associated to the objects. The intersection part is then excluded in the motion and shape estimation. Occlusion reasoning also improves the shape estimation in case of adjacent objects where shape estimates can be corrupted by image data of other objects. In this way we obtain robust motion estimates and trajectories for vehicles even in the case of occlusions, as we show in some experiments with real world traffic scenes.

1 Introduction and Related Work

Research in machine vision based traffic surveillance systems is of increasing interest for communities who suffer from high traffic density on highways. The task of a machine vision based traffic surveillance systems is to extract descriptions of moving vehicles from video data which enables further symbolic reasoning about the motion of the vehicles (i.e. [Koller *et al.* 91; Huang *et al.* 93]). In traffic scenes we have to cope with several moving objects which can interact with each other. Multitarget tracking requires not only an estimation algorithm that generates an estimate of the state of the object (target) to be tracked but also a data association component that decides which measurement to use for updating the state of which object. We assume our measurements to be corrupted by white Gaussian measurement noise (as in a single object tracking case) and by occlusions (measurements are missing or wrong due to overlayed data from other objects). We exploit a priori knowledge of the scene geometry in order to compute a depth order of the image regions associated to the moving objects and solve the data association problem by performing an explicit occlusion reasoning step.

Lecture Notes in Computer Science, Vol. 800
Jan-Olof Eklundh (Ed.)
Computer Vision – ECCV '94

Our tracker employs two linear Kalman Filters for estimating the control points of closed contours enclosing a moving region and the motion parameters according to an affine motion model. This tracker has been influenced by [Blake *et al.* 93], who successfully extended their real-time contour tracking system ([Curwen & Blake 92]) by exploiting affine motion models. They use so-called *Snakes* (active contour models) as descriptions of projected objects. Snakes have been introduced as deformable contours by [Kass *et al.* 88]. The Kalman-Filter theory ([Gelb 74]) is incorporated in the snake technique to form so-called *Kalman-Snakes* ([Terzopoulos & Szeliski 92]).

The common approach for updating a snake contour is to formulate an elastic-model approach, where forces are introduced based on image gradients normal to the contour. These approaches do not exploit the motion information contained in the images. Our approach deviates from this common snake technique in the sense that we do not use a force model, but we explicitly exploit motion information in the image if the object is moving.

There is little work on multi-object tracking applications using vision sensor data. [Meyer & Bouthemy 93] solve the problem of total occlusion by linking partial spatiotemporal trajectories using motion coherence, but they do not address the problem of data association and shifts in the estimated positions due to partial occlusions. Partial occlusions usually occur prior to a total occlusion and a missing occlusion reasoning step may already affect the shape and motion estimation which causes the motion coherence to be ineffective. [Létang *et al.* 93] realized that in the case of a partial occlusion the center of gravity of a blob mask is not a reliable feature to track. They handle the (partial) occlusion problem by tuning the measurement noise according to the change in blob size.

2 Motion Segmentation

An important component in tracking systems is *track* formation or *initialization*, for which we use a motion segmentation step. We use a modified version of the moving object segmentation method suggested by [Karmann & von Brandt 90] and implemented by [Kilger 92]. This method uses an adaptive background model, which is updated in a Kalman filter formalism, thus allowing for dynamics in the model as lighting conditions change. The background is updated each frame via the update equation

$$B_{t+1} = B_t + (\alpha_1(1 - M_t) + \alpha_2 M_t)D_t \tag{1}$$

Where B_t represents the background model at time t, D_t is the difference between the present frame and the background model, and M_t is the binary moving objects hypothesis mask. The gains α_1 and α_2 are based on an estimate of the rate of change of the background. The hypothesis mask, M_t, attempts to identify moving objects in the current frame. Our implementation differs from the one used in [Karmann & von Brandt 90; Kilger 92] in that we employ linear filters to increase the accuracy of the decision process. Thus in the notation above, B_t and D_t represent a vector of filtered responses instead of single images. We choose

choose as filter kernels the Gaussian and its derivative along the horizontal and along the vertical directions. An example of an identified moving object is shown in Figure 3b. For a complete description we refer the reader to [Koller *et al.* 93].

3 Contour Extraction and Shape Estimation

The contour extraction is based on motion and greyvalue boundaries. To extract candidate contour and motion points we simply threshold the spatial image gradient and the time derivative of the image function. A convex polygon enclosing all these sample points is then used as an initial object description. Figure 3 shows an image section with a moving car (a), the detected image patch covering the image of the car (b), and the image locations with well defined spatial gradient and time derivative constituting the sample points. The convex polygon enclosing all these sample points of c) is shown in d).

A convex polygon is not suitable for a time recursive shape estimation, since the number of vertices may and will change along the image sequence with new measurements. A solution is provided by *snakes*, spline approximation to contours [Kass *et al.* 88; Curwen & Blake 92]. We use closed cubic splines with 12 *control points* to approximate the extracted convex polygon. The locations of the control points are obtained by least squares between equidistant sample points along the contour of the polygon and the (uniform) spline segments ([Bartels *et al.* 87]). The details can be found in [Koller *et al.* 93]. The spline approximation of the contour for the previous example is given in Figure 3 e).

The state vector for the contour estimation along the image sequence comprises the estimates of the 12 pairs of x and y coordinates of the control points. As a measurement vector we use the the x and y coordinates of the control points obtained from the spline approximation of the extracted contour. This is the simplest case for the Kalman Filter where the measurement function is identical to the state vector itself ([Gelb 74]).

The support for the contour extraction is based on a binary image mask associated to a moving object. During the initialization this mask is identical to the moving image patch detected in the motion segmentation process. The predictions of the control vertices of the spline contour are then used to define the support in the tracking stage.

4 Motion Estimation

For a sufficiently small field of view and independently moving objects, the image velocity field $\boldsymbol{u}(\boldsymbol{x})$ at location $\boldsymbol{x}$ inside a moving image patch can be well approximated by a linear (affine) transformation. The degrees of freedom can be further reduced since the motion is constraint on a road plane and possible rotational components along the normal of the plane are small. We end up with a simple translation $\boldsymbol{u}_0$ and a change in scale s as the motion parameters and obtain:

$$\boldsymbol{u}(\boldsymbol{x}) = s\,(\boldsymbol{x} - \boldsymbol{x}_m) + \boldsymbol{u}_0, \tag{2}$$

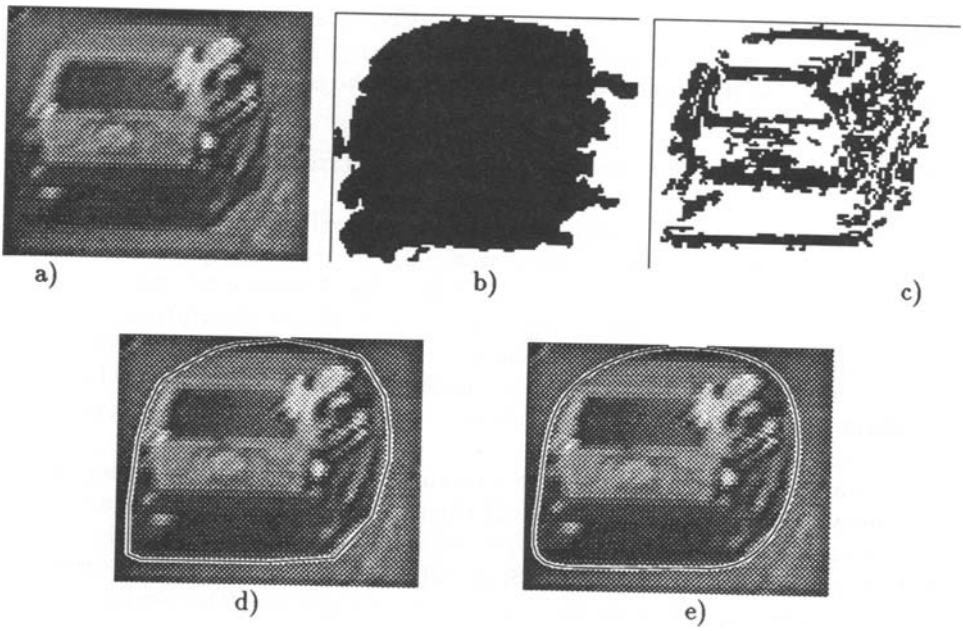

Fig. 1. a) An image section with a moving car. b) the moving object mask provided by the motion segmentation step. c) the image location with well defined spatial gradient and temporal derivative, used as sample points to define d), the convex polygon enclosing these points of c). e) the final contour description by cubic spline segments approximating the polygon of d).

$s = 0$ stands for no change in scale, while $s < 0$ and $s > 0$ stands for a motion component along the optical axes away and towards the camera, respectively.

The state vector for motion estimation comprises the affine motion parameters $\boldsymbol{\xi} = (\boldsymbol{u}, s)$. As a measurement we use also the coordinates of the control points of the spline contour extracted in a new acquired image. The measurement function can be expressed in a linear (matrix) equation and enables the use of a second linear Kalman Filter. Details can be found in [Koller *et al.* 93]. The initial value of $\boldsymbol{u}_0$ for an object is set to be the discrete time derivative of the objects center locations, measured in the first two frames. As an initial value for the scale parameter s we set $s = 0$.

5 Occlusion Reasoning

Any contour distortion due to partial occlusion will generate an artificial shift in the trajectory. In order to avoid these erroneous shifts and get reasonable tracks from the contours, we explicitly reason about occlusion. This is facilitated by

the special viewing geometry in our domain–the cars move on a ground plane. Nearer objects will appear lower in the image plane, and occlude farther away objects. This means that if we process the object contours starting from the bottom of the image plane, we can explicitly allow for the partial occlusion of the bounding contours of more distant objects [Koller *et al.* 93]).

The occlusion reasoning step also improves robustness in cases which we call *near* occlusion, where objects move very close next to each other so that the contours of the object will interfere and the estimation process can be confused by the presence of other object. Knowledge about the other contour provides the means to avoid image data from another object to be considered in the contour estimation of the object under investigation. For that purpose we perform the intersection analysis with about 5% enlarged contours in order to sense those cases. An example of tracking a car corrupted be an artificial occlusion is illustrated in Figure 2.

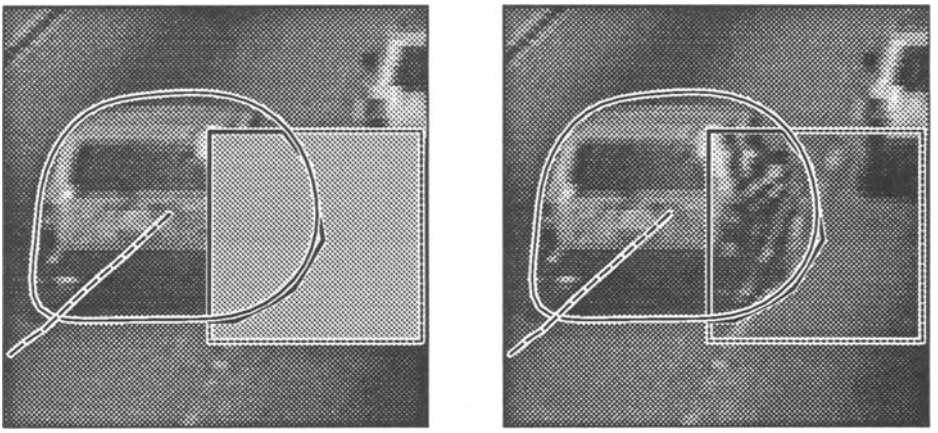

Fig. 2. An artificial occlusion (grey square) is stamped into images of an image sequence. The left image shows the occluding contour in thick lines, while the contour estimate of the object to be tracked is shown by thin lines. The vectors represent the trajectory of the object. The right image shows the lines overlayed to the original image without occlusion in order to compare the result.

6 Results with Real World Traffic Scenes

In order to validate our approach we conducted several experiments. We present here the result of tracking cars in an image sequence of 96 frames of a divided 4 lane freeway of the Los Angeles area. The left column of Figure 3 shows some intermediate images of the sequence with the overlayed contour estimates, while the right column shows the contour estimates with the tracks. In this sequence we have to cope with a partial occlusions and some near occlusions.

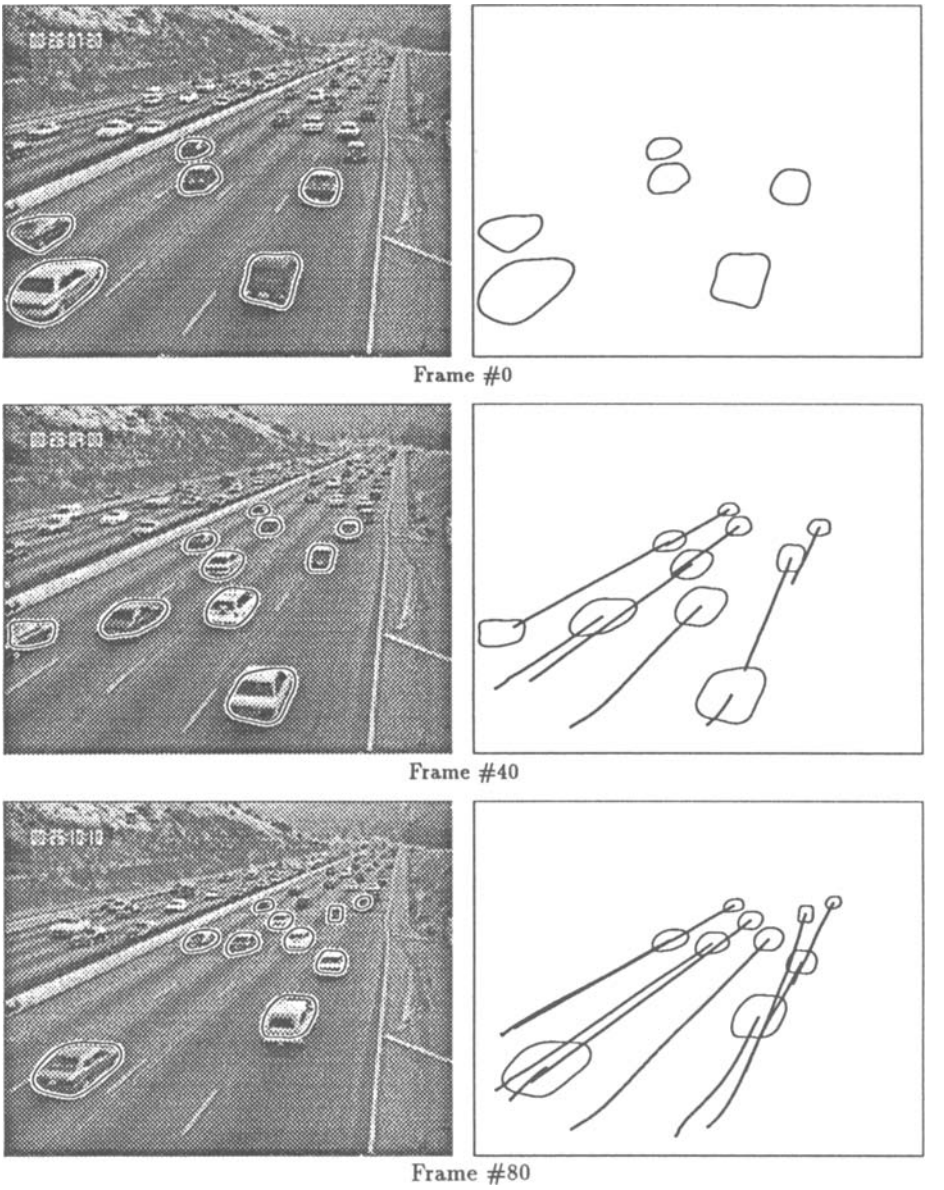

Fig. 3. The left column shows frame #0, #40, and #80 of the image sequence with the overlayed overlayed contour estimates of the cars. The right column shows the contour estimates with their tracks, starting from frame #0.

7 Conclusion

We designed a system for robust detection and tracking of multiple vehicles in road traffic scenes. The system provides tracks and shape description of vehicles which are suitable for further evaluation with symbolic reasoning in a traffic surveillance system. As typical for a surveillance system, the recording camera is assumed to be stationary, mounted, i.e., on a bridge or a pole beside the road, in order to cover a large field of view and to reduce occlusions of vehicles.

The objects are assumed to be well describable by convex contours and the motion is expected to be predominantly translational on a plane. We describe a contour by a closed cubic spline (known as *snakes*) and use an affine motion model for the innovation of the contour. The tracker is based on two simple Kalman Filters, estimating the affine motion parameters and the control points of the closed spline contour of the vehicles. As measurements we take the control points of spline contours approximating convex polygons, enclosing candidate motion and contour points.

The initialization of the tracker is performed by a kind of image differencing between a continuously updated background image and a newly acquired image. Update of the background image is based on the motion estimate. Trajectories of the moving vehicles are derived from the motion of the center of the control points of the closed spline contour.

In order to obtain reliable trajectories of vehicles in a highly cluttered environment — such as in highway scenes with heavy traffic — we have to solve the problem of data association for a multitarget tracking application. In the case of a partial occlusion, the center of the control points does not provide a reliable feature for the trajectory of an object, since the contour will be corrupted by wrong contour measurements. We solve this problem by an explicit occlusion reasoning step. Occlusion detection is performed by a depth ordered detection of overlapping contours associated to moving vehicles.

8 Acknowledgments

We gratefully acknowledge the help of C. McCarley and his group at at Cal Poly, San Luis Obispo in providing us with video tapes of traffic scenes covering all kind of traffic conditions. We also like to thank B. Rao for discussions about data association and multitarget tracking. This work was supported by California Department of Transportation through the PATH project grant no. MOU-83.

References

[Bartels *et al.* 87] R. Bartels, J. Beatty, B. Barsky, *An Introduction to Splines for use in Computer Vision*, Morgan Kaufmann, 1987.

[Blake *et al.* 93] A. Blake, R. Curwen, A. Zisserman, Affine-invariant contour tracking with automatic control of spatiotemporal scale, in *Proc. Int. Conf. on Computer Vision*, Berlin, Germany, May. 11-14, 1993, pp. 66–75.

[Cipolla & Blake 92] R. Cipolla, A. Blake, Surface Orientation and Time to Contact from Image Divergence and Deformation, in *Proc. Second European Conference on Computer Vision*, S. Margherita, Ligure, Italy, May 18-23, 1992, G. Sandini (ed.), Lecture Notes in Computer Science **588**, Springer-Verlag, Berlin, Heidelberg, New York, 1992, pp. 187–202.

[Curwen & Blake 92] R. Curwen, A. Blake, *Active Vision*, MIT Press, Cambridge, MA, 1992, chapter Dynamic Contours: Real-time Active Snakes, pp. 39–57.

[Gelb 74] A. Gelb (ed.), *Applied Optimal Estimation*, The MIT Press, Cambridge, MA and London, UK, 1974.

[Huang *et al.* 93] T. Huang, G. Ogasawara, S. Russell, Symbolic Traffic Scene Analysis Using Dynamic Belief Networks, in *AAAI Workshop on AI in IVHS*, Washington D.C., 1993.

[Karmann & von Brandt 90] Klaus-Peter Karmann, Achim von Brandt, Moving Object Recognition Using an Adaptive Background Memory, in V Cappellini (ed.), *Time-Varying Image Processing and Moving Object Recognition, 2*, Elsevier, Amsterdam, The Netherlands, 1990.

[Kass *et al.* 88] M. Kass, A. Witkin, D. Terzopoulos, Snakes: Active Contour Models, *International Journal of Computer Vision* **1** (1988) 321–331.

[Kilger 92] Michael Kilger, A Shadow Handler in a Video-based Real-time Traffic Monitoring System, in *IEEE Workshop on Applications of Computer Vision*, Palm Springs, CA, 1992, pp. 1060–1066.

[Koenderink 86] J.J. Koenderink, Optic flow, *Visual Research* **26** (1986) 161–180.

[Koller *et al.* 91] D. Koller, N. Heinze, H.-H. Nagel, Algorithmic Characterization of Vehicle Trajectories from Image Sequences by Motion Verbs, in *IEEE Conf. Computer Vision and Pattern Recognition*, Lahaina, Maui, Hawaii, June 3-6, 1991, pp. 90–95.

[Koller *et al.* 93] D. Koller, J. Weber, J. Malik, *Robust Multiple Car Tracking with Occlusion Reasoning*, technical report UCB/CSD-93-780, University of California at Berkeley, Oktober 1993.

[Létang *et al.* 93] J.M. Létang, V. Rebuffel, P. Bouthemy, Motion detection robust to pertubations: a statistical regularization and temporal integration framework, in *Proc. Int. Conf. on Computer Vision*, Berlin, Germany, May. 11-14, 1993, pp. 21–30.

[Meyer & Bouthemy 93] F. Meyer, P. Bouthemy, Exploiting the Temporal Coherence of Motion for Linking Partial Spatiotemporal Trajectories, in *IEEE Conf. Computer Vision and Pattern Recognition*, New York City, NY, June 15-17, 1993, pp. 746–747.

[Murray *et al.* 93] D.W. Murray, P.F. McLauchlan, I.D. Reid, P.M. Sharkey, Reactions to Peripheral Image Motion using a Head/Eye Platform, in *Proc. Int. Conf. on Computer Vision*, Berlin, Germany, May. 11-14, 1993, pp. 403–411.

[Rao 92] B.S.Y. Rao, *Active Vision*, MIT Press, Cambridge, MA, 1992, chapter Data Association Methods for Tracking Systems, pp. 91–105.

[Terzopoulos & Szeliski 92] D. Terzopoulos, R. Szeliski, *Active Vision*, MIT Press, Cambridge, MA, 1992, chapter Tracking with Kalman Snakes, pp. 3–20.

[Zheng & Chellappa 93] Q. Zheng, R. Chellappa, Automatic Feature Point Extraction and Tracking in Image Sequences for Unknown Camera Motion, in *Proc. Int. Conf. on Computer Vision*, Berlin, Germany, May. 11-14, 1993, pp. 335–339.