On the Classification of Moving Objects in Image Sequences Using 3D Adaptive Recursive Tracking Filters and Neural Networks

L. T. Bruton[1], *Fellow, IEEE*, N. R. Bartley[1], *Member, IEEE*, and
Z. Q. Liu[2], *Senior Member, IEEE*

[1]Department of Electrical and Computer Engineering, University of Calgary, Calgary, Alberta, Canada.
[2]Department of Computer Science, University of Melbourne, Parkville, Victoria, Australia.

Abstract

It is shown that 3D recursive filters may be used to classify the motion of objects in discrete-time spatio-temporal 3D image sequences. The 3D filter has a time-varying 3D frequency-planar passband that is adapted in a feedback system to automatically track a moving object on the basis of its smoothly changing trajectory, thereby rejecting noise and stopband objects that are not of interest. The adaptive space-time velocity vector of the passband object is available within this feedback system and is used as the input to a multi-layer perceptron neural network which classifies the motion of the passband object according to a number of motion characteristics, such as its direction of travel, velocity, acceleration as a function of time and position and its stopping time. It is shown that such a system may be used to classify the motion of vehicles at an intersection of roads.

1 Introduction

Three dimensional (3D) recursive digital Linear-Trajectory (LT) filters may be used to selectively track the motion of objects in digital image sequences by adaptively steering the 3D frequency-planar passband of the filter to track the slowly time-varying 3D frequency-planar energy density function of the moving object in space-time [1]. Such tracking methods are known [1] to be especially immune to additive noise and they can successfully discriminate against objects that have motions that differ only slightly from that of the object that is being tracked.

The purpose of this contribution is to show that the selectively filtered tracked object, which is a passband object of the LT filter, produces a robust estimate of the instantaneous 3D space-time velocity vector of the tracked object which may then be used as a member of the batched training input matrix of a conventional Multilayer Perceptron (MLP) neural network [2]. The outputs of the MLP can be used as classifiers that identify and categorize many useful motion and positional characteristics of the objects in the image sequence.

The results reported here use a first-order 3D recursive filter [1],[3] and a two-layer 100-input 4-output MLP to correctly classify vehicles that have 4 distinct trajectories as they pass through a two-road intersection in a digitized image sequence. It is shown that classifications on the basis of vehicle velocities and directions of travel are feasible with 100% classification accuracy.

The selectivity of the 3D passband of the filter determines the ability of the filter to discriminate against competing objects and noise in the input sequence. The results suggest that the method will be sufficiently robust to classify the motions of the tracked vehicles when the input image sequences are heavily contaminated with additive noise or with other objects that enter the image having trajectories in space-time that are close to the object that is being tracked. For example, for the case of vehicle motion classifications, the method produces motion/position classifications that are relatively robust in the presence of other vehicles, pedestrians and video noise.

2 Review of 3D linear trajectory signals and filters

In this section, we briefly review the fundamental class of 3D LT signals and the simple 3D recursive filter algorithms that have been used to enhance LT signals.

2.1 3D linear trajectory signals

Many important 3D filtering problems, associated with the enhancement of digitized image sequences, involve 3D component signals that can be characterized in the 3D continuous space-time domain $\mathbf{t} \equiv (t_1, t_2, t_3) \in \Re^3$ as LT signals. A LT signal $v_{LT}(\mathbf{t})$ is defined in [3] as belonging to the class of 3D signals for which there exists a direction $\mathbf{d}$ along which $v_{LT}(\mathbf{t})$ is constant along all lines in this direction, as illustrated in Fig. 1(a). When presented as a temporal succession of frames in t_3, such signals have the property that they move with constant spatial velocity in t_1, t_2 [3].

The 3D LT signal in Fig. 1(a) has a corresponding energy density spectrum that exists entirely on a plane passing through the origin in $\omega \equiv (\omega_1, \omega_2, \omega_3) \in \Re^3$, as shown in Fig. 1(b). The normal $\mathbf{n}$ to this plane, referred to as the *signal plane*, has the same direction in ω as the direction $\mathbf{d}$ of the LT signal in $\mathbf{t}$ [3].

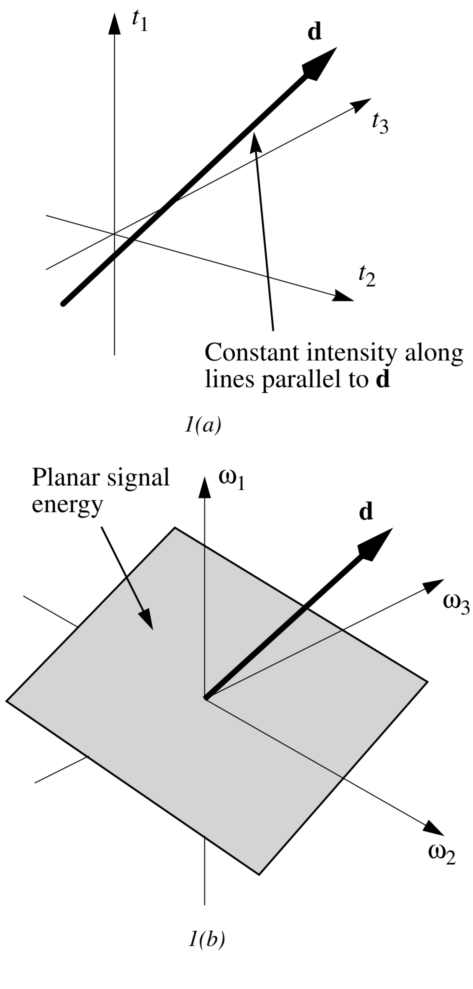

Figure 1: 3D LT signals

2.2 3D recursive linear trajectory filters

Simple 3D continuous-domain prototype filter networks, comprised of inductance (L), capacitance (C) and resistance (R) elements, may be designed to have a passband that closely surrounds the signal plane and have been shown to be useful for selectively enhancing such LT signals [3],[4]. One such network that has been employed successfully [3],[4] is shown in Fig. 2.

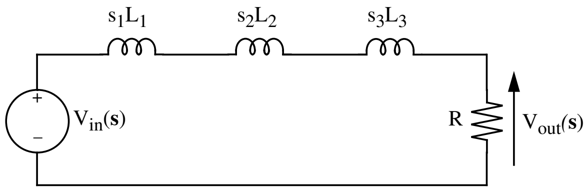

Figure 2: A 3D LT prototype filter network

The Laplace transform input-output voltage transfer function $T(\mathbf{s}), \mathbf{s} \equiv [s_1, s_2, s_3]^T \in \mathcal{C}^3$, is given by

$$T(\mathbf{s}) = \frac{R}{R + s_1 L_1 + s_2 L_2 + s_3 L_3} \quad (1)$$

and has a 3D magnitude frequency response $M(\omega) = |T(j\omega)|$ given by

$$M(\omega) = \frac{R}{\sqrt{R^2 + (\omega_1 L_1 + \omega_2 L_2 + \omega_3 L_3)^2}}. \quad (2)$$

$M(\omega)$ has a resonant plane, as shown in Fig. 3, and given by

$$\omega_1 L_1 + \omega_2 L_2 + \omega_3 L_3 = 0 \quad \textbf{Resonant Plane} \quad (3)$$

and has -3dB planes, also shown in Fig. 3, given by

$$\omega_1 L_1 + \omega_2 L_2 + \omega_3 L_3 = \pm R. \quad \mathbf{-3dB\ Planes} \quad (4)$$

The resonant plane in Fig. 3 passes through the origin and has unit normals $\mathbf{n}$ given by $\mathbf{n} = \pm ||L||_2^{-1}(e_1 L_1 + e_2 L_2 + e_3 L_3), \quad L_1, L_2, L_3 \geq 0$, where $||L||_2 = \sqrt{L_1^2 + L_2^2 + L_3^2}$ and e_1, e_2, and e_3 are the orthogonal unit basis vectors in the $\omega_1, \omega_2, \omega_3$ directions. The perpendicular distance between the resonant and -3dB planes is $R/||L||_2$ and the 3D Q-factor for this filter is defined in [3] as $Q(\omega) = ||\omega||_2 ||L||_2 / R$.

Corresponding discrete-domain LT image sequences $v_{LT}(x, y, t), x, y, t \in \mathcal{N}^1$, may be obtained from $v_{LT}(\mathbf{t})$ by uniformly sampling. We employ the triple bilinear transformation on (1), which leads to simple 3D recursive filter algorithms for the enhancement of 3D digitized LT image sequences [3],[4].

3 Review of the adaptive 3D recursive linear trajectory filter system

In this section, we briefly review the adaptive 3D recursive LT filter described in [1] and show how object velocity information is extracted for the purpose of motion classification.

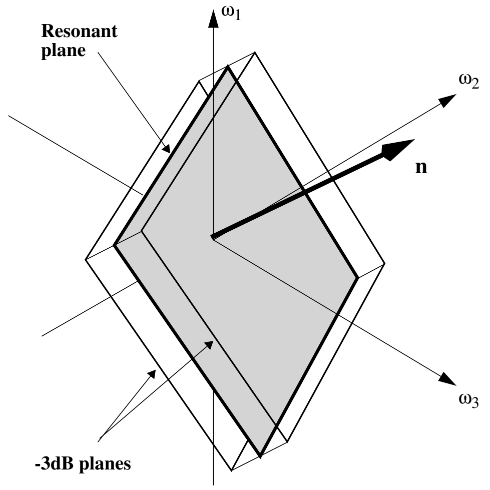

Figure 3: The resonant and -3dB planes of the 3D LT filter

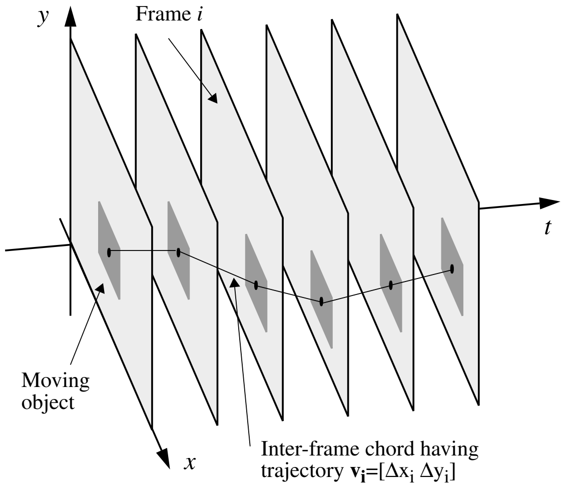

Figure 4: A piecewise linear trajectory signal

In many practical 3D image sequences, objects often do not proceed on perfectly linear trajectories, but instead exhibit temporally-smooth motion along curved trajectories in space-time. It is convenient to represent such curved-trajectory signals in x, y, t with piecewise linear trajectory signals, where the trajectory is shown in Fig. 4 as a time sequence of chordal segments. The enhancement of such piecewise linear trajectory signals is then carried out by employing an *adaptive* 3D LT filter whose resonant plane is oriented on every frame i to selectively enhance an LT signal having the direction of the chordal segment $\mathbf{v_i}$, as shown.

We employ the adaptive 3D filter system shown in Fig. 5. Consider the output of the 3D LT filter at frame $t = i$ and let the position of the object be centered at the pixel position $\bar{x}_i, \bar{y}_i$, and similarly at $\bar{x}_{i-1}, \bar{y}_{i-1}$ in frame $i-1$. We define the instantaneous velocity vector as $\mathbf{v_i}$ where

$$\mathbf{v_i} = [\Delta\bar{x}_i \;\; \Delta\bar{y}_i] = [\bar{x}_i - \bar{x}_{i-1} \;\; \bar{y}_i - \bar{y}_{i-1}]. \qquad (5)$$

We use $\mathbf{v_i}$ to define the trajectory of the object on frame i.

Following [1], we determine the position of the object in the output of the 3D LT filter on each frame by computing the intensity centroid within an object window, as shown in Fig. 6. For a window size of X_W, Y_W, located at X_i, Y_i, the centroid of the object $\bar{x}_i$, $\bar{y}_i$, is given by

$$\bar{x}_i = \frac{1}{A}\sum_{x=0}^{X_W}\sum_{y=0}^{Y_W} V_{out}(x + X_i, y + Y_i, i)x$$

$$\bar{y}_i = \frac{1}{A}\sum_{x=0}^{X_W}\sum_{y=0}^{Y_W} V_{out}(x + X_i, y + Y_i, i)y \qquad (6)$$

$$A = \sum_{x=0}^{X_W}\sum_{y=0}^{Y_W} V_{out}(x + X_i, y + Y_i, i)$$

The object's centroid $\bar{x}_i, \bar{y}_i$ is then used as the input to the feedback portion of the system in Fig. 5, where the instantaneous time-varying velocity estimate $\mathbf{v_i}$ is determined. This estimate is then filtered using two second-order 1D lowpass recursive filters to smooth the frame-to-frame discontinuities that occur in the velocity estimates. Finally, based on the smoothed velocity estimates $\hat{\mathbf{v}}_\mathbf{i} = [\hat{x}_i \;\; \hat{y}_i]$, the time-varying coefficients b_{jkl} of the 3D LT filter are determined for each frame i. We propose here *to employ this smoothed velocity estimate* $\hat{\mathbf{v}}_\mathbf{i}$ *directly to classify the trajectory of the object*, as described in the following section.

4 Trajectory classification of the passband tracked object

For the purpose of classifying the trajectory of a passband object, tracked by the above 3D tracking system, we propose here to employ a conventional static MLP network and Time-Delay Neural Network (TDNN) [2], as shown in Fig. 7.

The TDNN consists of a tapped delay line of length M having unity tap weights and is used to collect a time series of M velocity estimates $\hat{\mathbf{v}}_\mathbf{i}$. Velocity estimates are input to the TDNN when the 3D tracking system has captured a particular object. After M estimates have been collected, the TDNN output vec-

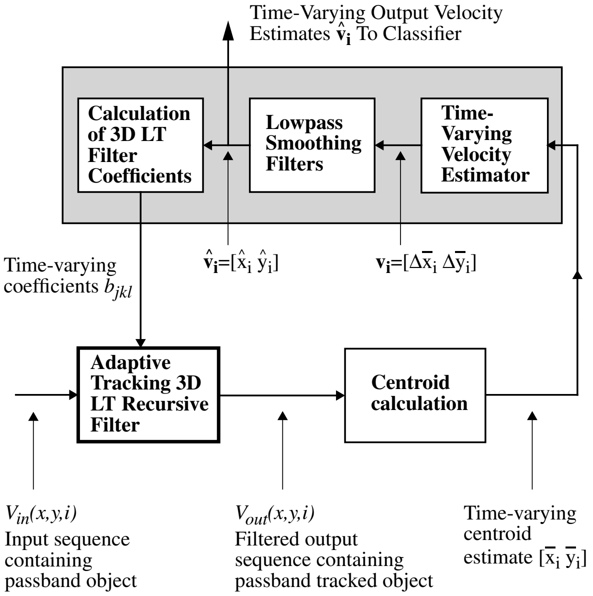

Figure 5: The 3D system for tracking a moving object and obtaining velocity estimates

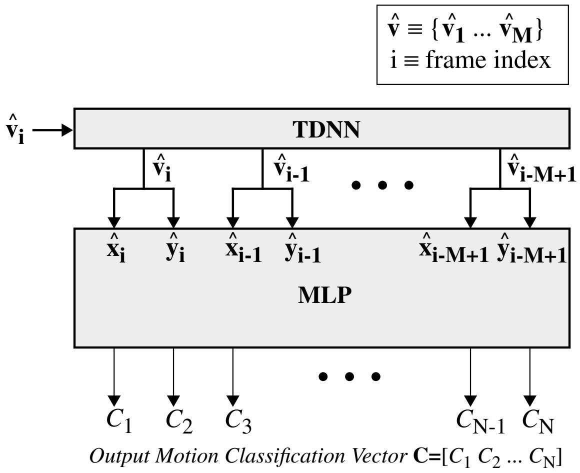

Figure 7: The neural network used to classify N motion trajectories using M velocity estimates

tor, which represents the *static velocity profile* of the object, is used as the input to the MLP to achieve classification.

The MLP is a conventional two-layer network with $2M$ inputs, N outputs and employs log-sigmoid transfer functions in each layer. Each of the N outputs corresponds to a particular velocity profile to which the network has been trained to classify. Training of the network is achieved by applying a matrix of N such velocity profiles in batch to the network and by employing the conventional back-propagation training algorithm [2].

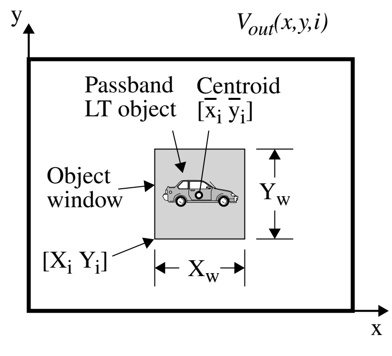

Figure 6: The intensity centroid of the passband LT object on frame i

5 Classification of moving vehicles

We have employed the above tracking and classification algorithm to classify the trajectories of vehicles passing through an intersection. A 10-second video sequence has been obtained and contains vehicles that move on four distinct trajectories. The trajectories correspond to vehicles that pass directly through the intersection at constant speed in the left and right directions and vehicles that begin left and right turns, pausing to yield to oncoming traffic. The trajectories are shown in Fig. 8 and have been assigned the numeric classifications 0-3, as shown. A 100-input 2-layer MLP has been trained to classify the velocity profiles of each of these four trajectories.

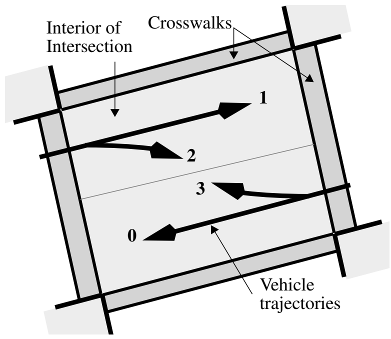

Figure 8: Vehicle trajectory classifications

The detection of vehicles for tracking and classification has been automated by placing a number of fixed *sensor windows* on each lane of traffic where vehicles enter the intersection, as shown in Fig. 9. In each sensor window, we employ a fixed 3D LT filter having a passband that is tuned to enhance vehicles moving in the direction and average speed of traffic for that lane. The selectivity of the passband is less than that

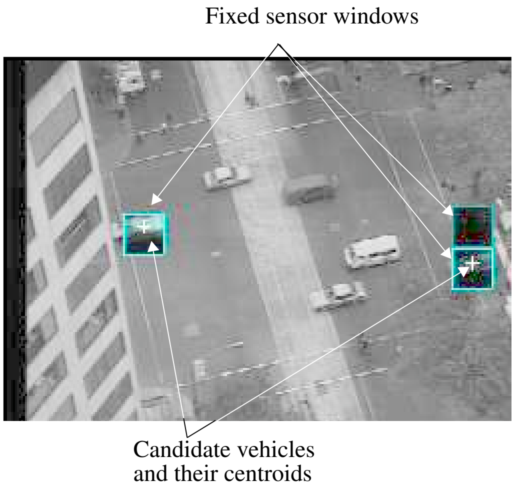

Figure 9: Vehicle detection using sensor windows

used in the tracking algorithm so that vehicles that are slightly slower or faster than the average speed are still transmitted by the filter. The total output intensity A in equation (6) for each sensor window is determined and monitored on each frame. Once A has exceeded a certain threshold in a particular window, the vehicle in the window becomes a candidate for tracking. As shown in Fig. 9, two such candidates have been identified, as indicated by the brighter window borders and cross-hairs over each vehicle's centroid. After the candidate-vehicle intensity A has peaked in the window, the current position and velocity estimates are used to initialize the tracking/classification algorithm for that vehicle.

The overall system of vehicle detection, tracking and classification is shown at work in Fig. 10. Each vehicle that is being tracked is indicated by its surrounding rectangular object window containing the LT-filtered output image and cross-hairs over the centroid. For vehicles that have been tracked for 50 frames, the 50 velocity estimates $\hat{\mathbf{v}}_{\mathbf{i}} = [\hat{x}_i \ \hat{y}_i]$ are input to the MLP for classification. The resulting numeric classification is then "superimposed" on the roof of the vehicle, as shown for two such vehicles in Fig. 10 which are classified as "0" and "1". Although not shown here, 100% classification accuracy has been achieved for all tracked vehicles in the sequence.

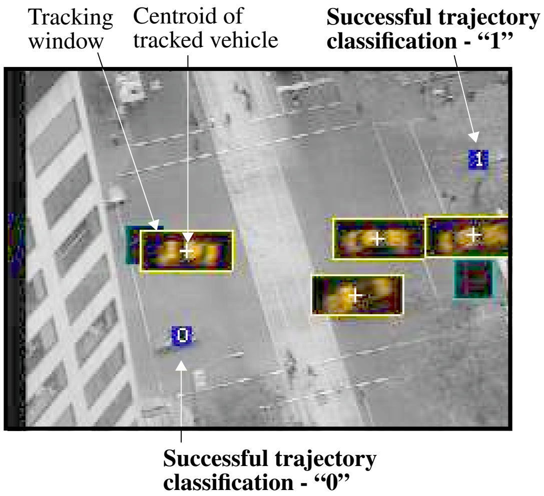

Figure 10: Vehicle tracking and classification

6 Summary

We have described a new application of 3D space-time filters and neural networks for the purpose of classifying the motion of objects in image sequences, and have applied it to the particular case of vehicle motion classification. An adaptive 3D LT recursive tracking filter may be used to obtain a robust estimate of the velocity of a particular vehicle in the sequence. Accurate classification is achieved using a conventional MLP neural network.

Acknowledgements

The authors gratefully acknowledge the financial support of Micronet, the Canadian Federal Centre of Excellence on Microelectronic Devices, Circuits and Systems and the Natural Sciences and Engineering Research Council of Canada.

References

[1] L. T. Bruton and N. R. Bartley, "The enhancement and tracking of moving objects in digital images using adaptive three-dimensional recursive filters", *IEEE Trans. on Circuits and Systems*, vol. CAS-33, pp. 604–612, June 1986.

[2] D. R. Hush and W. G. Horne, "Progress in supervised neural networks", *IEEE Signal Processing Magazine*, vol. 10, pp. 8–39, January 1993.

[3] L. T. Bruton and N. R. Bartley, "Three-dimensional image processing using the concept of network resonance", *IEEE Trans. on Circuits and Systems*, vol. CAS-32, pp. 664–672, July 1985.

[4] Y. Zhang and L. T. Bruton, "Applications of 3-D LCR networks in the design of 3-D recursive filters for processing image sequences", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. CSVT-4, pp. 369–382, August 1994.