SAND2015-10245C

Vehicle track detection in CCD imagery via conditional random field

Rebecca Malinas Tu-Thach Quach Mark W. Koch
Sandia National Laboratories*
Albuquerque, NM 87185-1163
rmalina@sandia.gov, tong@sandia.gov, mwkoch@sandia.gov

Abstract

Coherent change detection (CCD) can be used to detect subtle scene changes in synthetic aperture radar (SAR) imagery, such as vehicle tracks, which indicate human activity. Automatic track detection in SAR CCD is difficult due to various sources of low coherence other than the track activity we wish to detect. Existing methods require user cues or explicit modeling of track structure, which limit algorithms' ability to find tracks that do not fit the model. In this paper, we present a track detection approach based on a pixel-level labeling of the image via a conditional random field classifier, with features based on radial derivatives of local Radon transforms. Our approach requires no modeling of track characteristics and no user input, other than a training phase for the unary cost of the conditional random field. Experiments show that our method can successfully detect both parallel and single tracks in SAR CCD as well as correctly declare when no tracks are present.

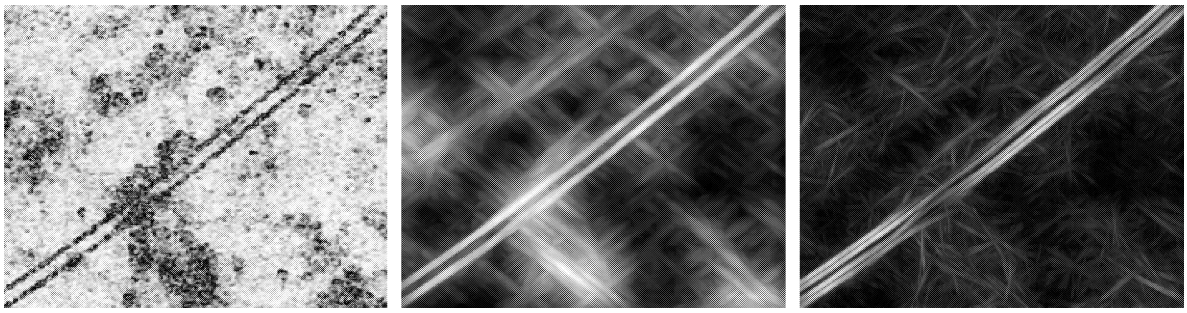
1. Introduction

Synthetic aperture radar (SAR) coherent change detection (CCD) can indicate subtle scene changes, such as vehicle tracks left by an individual driving through the scene [7]. Detection of these tracks is useful for surveillance and search and rescue applications; however, automatic detection of vehicle tracks in SAR CCD is difficult due to various sources of low coherence other than the vehicle track change we wish to detect, such as ground surface change due to weather effects and vegetation, registration errors, and radar shadows. Current techniques for vehicle track detection in SAR CCD [4, 5] require user cues for identification of roads or other indicators for possible vehicle activity, or explicit modeling of vehicle tracks as parallel low coherence areas. This knowledge is incorporated into a template that is used to detect candidate vehicle track points, followed by a curve-fitting step to link points. The parallel template causes difficulty in detecting single tracks. Another method [10] detects tracks by first identifying likely track points, and then selecting the simplest model for a set of curves that best fits the data. These methods may also involve a pre-processing smoothing or thresholding step to mitigate false alarms caused by the non-track change described previously, and some methods can find only a single track in an image. In this paper, we present a pixel-level labeling (track or non-track) of CCD images based on a conditional random field classifier. We make no assumptions about vehicle track characteristics, other than the general assumption that man-made tracks are long and thin in comparison to other CCD clutter. We require no user input and no pre-processing steps, other than a training phase for the unary cost of the conditional random field. Since we do not explicitly model parallel vehicle tracks, this method has application to detecting other change phenomena, such as single vehicle tracks, building edges, or ground surface disturbances. Our unary cost is derived from radial derivatives of local Radon transforms centered on each pixel to be labeled. Experimental results show that our method successfully detects both parallel and single tracks in SAR CCD imagery and can also correctly label pixels in clutter areas as non-track when no tracks are present.

2. Technical Approach

We frame track detection as a binary labeling problem on a CCD image: we classify each image pixel as belonging to a track or not. Associated with each pixel n is observation x_n and true class y_n. We model conditional dependencies between the class variables $\mathbf{y}$ as

*Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.



(a) (b) (c)

Figure 1: RT vs. RDRT: a) Example CCD image with track, b) Maximum values of RT windows, c) Maximum values of RDRT windows. Note that the derivative of the RT suppresses background clutter.

a Markov random field, and thus we model the posterior probabilities $p(\hat{\boldsymbol{y}}|\boldsymbol{x})$, for labels $\hat{\boldsymbol{y}}$, as a conditional random field. Given a set of N image pixels, we seek the image labeling $\hat{\boldsymbol{y}}$ such that [9]

$$\hat{\boldsymbol{y}} = \underset{y_1,\ldots,y_N}{\operatorname{argmin}}\Big\{ \sum_{n=1}^{N} (U_n(y_n) + \sum_{m\in\mathcal{N}(n)} P_{m,n}(y_m, y_n))\Big\}, \quad (1)$$

where $y_n \in \{0, 1\}$ is the class assignment of pixel n, $U_n(y_n)$ is the unary cost associated with assigning label y_n to the observed (feature) value at pixel n, and $P_{m,n}(m, n)$ is the pairwise cost associated with labeling neighbors m and n as y_m and y_n, respectively. Neighbors are determined using a 4-neighbor system.

2.1. Radial Derivative of the Radon Transform

The Radon transform (RT) of a continuous 2D image $f(x, y)$ is given by

$$\mathcal{R}(r, \theta) = \int_t f(t\sin\theta + r\cos\theta, -t\cos\theta + r\sin\theta)dt. \quad (2)$$

At (r, θ) in the transformed space, the RT takes the value of a line integral along the ray in $\mathbb{R}^2$ such that the closest point of the ray to the origin can be represented in polar coordinates as (r, θ). Given an image for labeling, we base our unary cost for a pixel on the radial (along r) derivative of a local Radon transform (RDRT) centered at that pixel. The RDRT has shown good performance in detection of faint tracks in SAR imagery [6], and also effectively distinguishes tracks from other background clutter in CCD imagery. To motivate the use of the RDRT instead of the RT for distinguishing tracks from background clutter, consider Figure 1. Figure 1 (a) shows a CCD image that contains large areas of contiguous clutter in addition to tracks. Figure 1 (b) is the result of computing RTs over 32x32 windows centered on each pixel in the original image, and then taking the maximum value from each window, *i.e.*, each pixel in (b) is the maximum of a RT computed over a window centered at that pixel. Details about this process are explained in a later section. Figure 1 (c) gives the maximum values of the RDRT from each window. For image (a), as shown in Figure 1 (b), both background clutter and tracks have large RT values. Hence, the RT value itself may not discriminate tracks from background clutter. Now we consider the rate-of-change of RT values for thin tracks vs. background clutter. For thin tracks, at a given θ, the derivative of the RT will have large values only at r that coincide with the track edges. On the other hand, clutter will result in a range of r with similar RT values, and so clutter will generally result in low derivatives of the RT, as shown in Figure 1 (c). Although there may be high derivatives at edges of clutter regions, the magnitude of the derivatives at clutter edges are generally lower than that of track edges, as visible in Figure 1 (c). Hence, the RDRT has the potential to distinguish tracks from background clutter, and we take advantage of this fact in our labeling.

2.2. Track Detection via Conditional Random Field

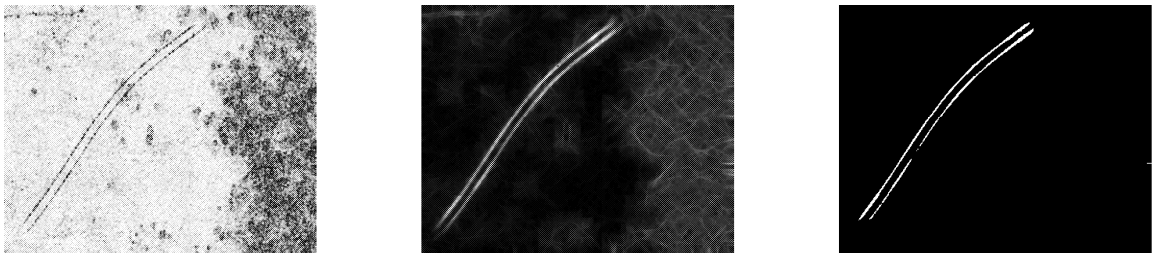
The unary cost $U_n(y_n)$ at pixel $n \in 1, ..., N$ takes the form

$$U_n(y_n) = -log(Pr(x_n|y_n)), \quad (3)$$

where x_n is the maximum RDRT value over a window centered on pixel n. The probabilities $Pr(x_n|y_n)$ are determined using logistic regression on ground truth images containing tracks. We choose the submodular pairwise cost

$$P_{m,n}(y_m, y_n) = \begin{cases} 1 & y_m \neq y_n \\ 0 & y_m = y_n \end{cases} \quad (4)$$

for neighboring (in a 4-neighbor system) pixels m, n. The magnitudes of the pairwise costs were chosen to have value at most 1 to encourage continuity along tracks while preserving track edges.



(a) (b) (c)

Figure 2: Example RDRT and labeling: a) Example CCD image with track, b) Maximum values of RDRT windows, c) Labeling.

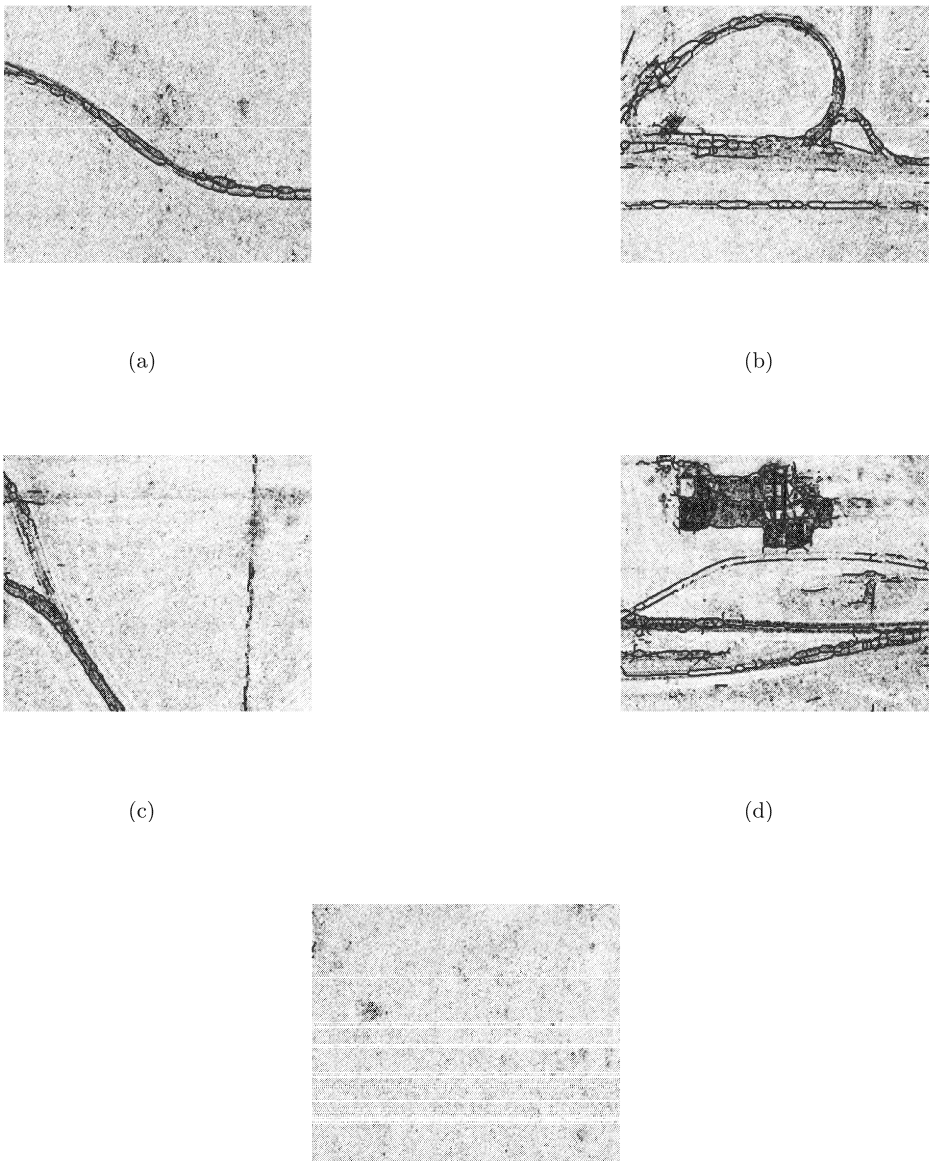
3. Implementation and Experimental Results

3.1. Implementation details

We compute local RTs over 32x32 windows using the sliding window method in [6]. If an image contains N pixels, we compute N RTs, where each RT window is centered on one of the image pixels (and hence any two adjacent windows will have 31 rows or columns of overlap). This window size was chosen as the optimal tradeoff between separation in the track pixel and non-track pixel RDRT value distributions and computation time. To obtain valid RT values at edge pixels, we pad the CCD image by replicating edges. After computing an RT for each image pixel (*i.e.*, centered on each image pixel), we find the numerical gradient of the local RTs along the r-dimension. To find a single value at each pixel for input to the unary function, we take the maximum magnitude of the RDRT in each window (hence, we obtain N maximum RDRT values, one for each pixel). These are the values x_n in (3). We find the optimal labeling $\mathbf{y}$ via graph cuts [2, 3, 8]. Figure 2 shows an example CCD image containing tracks, the maximum RDRTs, and the resulting conditional random field labeling. Due to high gradients at the boundaries of isolated clutter or buildings, we re-label any small connected components of track-labeled pixels as non-track. A good threshold for this re-labeling can be computed from the size distribution of connected components over the detector outputs of a set of test images.

3.2. Experimental Results

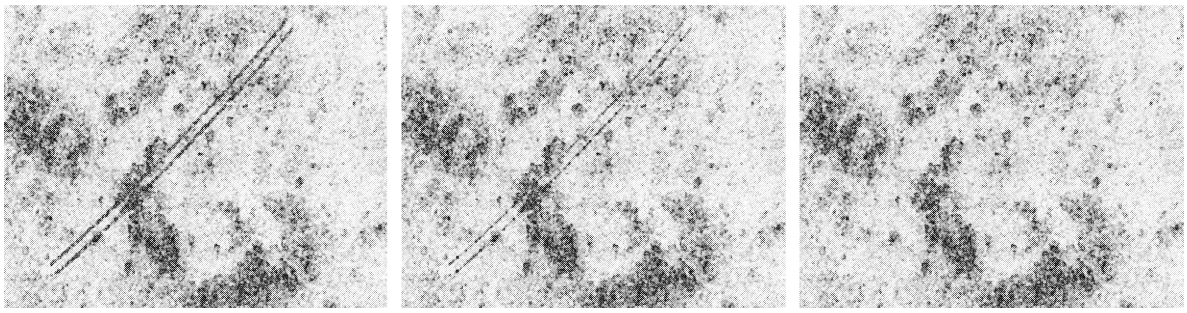
Figure 3 shows results of our algorithm on real SAR CCD images from a publicly available dataset [1]. There has been no removal of small components in these images. Note that Figure 3 (c) shows detection of a single-track-like structure. In Figure 3 (e) the CRF correctly identifies no tracks (the small detections at the image edges are caused by the padding step for the Radon transform windows, and can be removed by the thresholding step described previously). Due to the lack of available ground truth in collected data, we quantify the performance of our algorithm on a set of 40 real CCD images of size 600x800 pixels that contain simulated tire tracks of various curvatures. Each CCD image is generated from a real SAR image pair (from the same publicly-available data set) and contains a single randomly-generated pair of tire tracks. For a given SAR image pair of the same scene and randomly-generated pair of tire tracks, we place the tire tracks in the output CCD image by adding random (Gaussian) phase shifts to pixels along the track trajectory in the non-reference image of the SAR image pair [12]. We generate two test sets of varying thicknesses: we call the two sets *dark* and *medium*. Example images from each of the two sets as well as a CCD background image containing no tracks are shown in Figure 4. We use the same set of CCD background images and simulated tracks to generate both test sets for fair comparison among the two sets. This allows comparison of the algorithm results in response to different signal-to-noise ratios. The average track length is 705 pixels. We also tested our algorithm on the CCD background images, with no simulated tracks added. The algorithm correctly labeled all pixels as non-track in all but three of the images; in these three images, some linear clutter along the image edges were labeled as track, due again to the padding in the Radon transform step. These components can again be removed by thresholding on component size. In quantifying the performance of our algorithm, ground truth is taken to be the midline of the tire tracks (midline between the two tracks). Detection is quantified with respect to this midline. To compare candidate track pixels (labeled as a track pixel by the algorithm) to ground truth, we compute the



(a) (b) (c) (d) (e)

Figure 3: Results on real SAR CCD images. No post-processing or removal of small components has been performed on these images. (c) shows detection of a single-track-like structure in addition to parallel tracks. In (d), the CRF picks up building edges in addition to tracks. Approaches to mitigate detection of buildings is discussed in a later section. (e) contains no tracks, and the CRF correctly declares that. The small detections at the left edge of the image are caused by the padding step for the Radon transform, and can be removed through thresholding on component size.

tire track midline via morphological dilation followed by erosion. In the following, we define $D(p_1, p_2)$ as the Euclidean distance (in pixel space) between pixels p_1 and p_2. Let m denote a midline pixel we wish to de-



(a) (b) (c)

Figure 4: Example test images: (a) *dark*, (b) *medium*, (c) original CCD image. Each set consists of 40 images of size 800×600 containing various track curvatures and background clutter. The average track length is 705 pixels.

tect, d the candidate track pixel closest in Euclidean pixel distance to m, and t the closest tire track pixel to d. Then m is detected if and only if d is contained within the area bounded by the outer edges of the tire tracks, *i.e.*, candidate pixels may lie on the tracks or between the tracks. We write this mathematically as the condition

$$D(m, d) \leq D(m, t). \tag{5}$$

To evalute false alarms, for a given candidate track pixel d, we define m as the track midline pixel closest to d, and t as the tire track pixel closest to d. Then d is a false alarm if and only if d lies outside the tire track area (detections may lie between the tracks). We write this mathematically as the condition

$$D(m, d) > D(m, t). \tag{6}$$

Given the above definitions of detection and false alarms, we quantify the accuracy of our algorithm using the following metric:

$$\frac{CD}{GT + FA}, \tag{7}$$

where GT denotes the number of ground truth midline pixels, CD denotes the number of correct detection pixels, and FA denotes the number of false alarm pixels. Our metric gives a real number between 0 and 1, with value 1 if and only if all ground truth pixels are correctly detected with zero false alarms. The metric value will decrease as the number of false alarm pixels increases or the number of correct detections decreases. We note that this metric only quantifies the tradeoff between correct detections and false alarms, and does not consider number of tracks in the image or the location/distribution of false alarms in the image. For an image with multiple true tracks, the metric would not quantify the number of tracks found. A metric to quantify these aspects of track detection remains an open problem.

Figure 5 (a) shows the mean, median, and standard deviation of the metric values for each of the two test sets. We use the same unary and pairwise costs in the labeling of both test sets. The results show that the algorithm performs well for the *dark* and *medium* tracks. Previous work [10] using the Bayesian Information Criterion (BIC) for detection achieved similar results; however, our method can detect single tracks. Figure 5 (b) shows results using the BIC method, for comparison. Our method shows slightly increased mean and median accuracies for the *dark* and *medium* tracks, with comparable or lower standard deviation.

4. Discussion

SAR CCD can be used to detect vehicle tracks and other subtle scene changes; however, automatic track detection is difficult due to a variety of sources of low coherence. Existing methods often require user cues or modeling of vehicle tracks which limit performance. Our approach successfully detects both single and parallel vehicle tracks in CCD images, with only a training phase for the data cost of the conditional random field. Our algorithm does not assume a particular track structure, and thus could be applied to detection of other human activity. The algorithm showed good performance for typical tracks. In future work, we will address two issues: detection of building edges and improved detection of very faint tracks. To mitigate detection of building edges, we can use a post-processing step that distinguishes label configurations of tracks from those of buildings, or we could first detect buildings (using a pre-existing building detector, such as [11]) in images and ignore track detections that occur on top of the building detections. Alternatively, we could add a feature to the CRF that

	Mean	Median	Std. Dev.
dark	0.9840	0.9973	0.0562
medium	0.9015	0.9418	0.1009

(a) Conditional random field

	Mean	Median	Std. Dev.
dark	0.9721	0.9872	0.0471
medium	0.8352	0.9108	0.1812

(b) Bayesian Information Criterion

Figure 5: (a) Mean, median, and standard deviation of accuracies for the three test sets. (b) Previous results from [10].

discriminates tracks from buildings. The detection of very faint tracks can be improved using larger (64x64) Radon transform windows; however, the use of larger windows also increases false alarms. To both improve detection while keeping a low false alarm rate, a multiscale approach may be appropriate, in which we fuse detector outputs from multiple Radon window sizes.

Acknowledgment

This work was supported by PANTHER, a Laboratory Directed Research and Development (LDRD) Project at Sandia National Laboratories. For additional information about PANTHER, please contact Kristina Czuchlewski, Ph.D., krczuch@sandia.gov.

References

[1] W. J. Bow, Jr. Sandia sar data collect 2006. Technical Report Technical Report SAND2006-2290P, Sandia National Laboratories, 2006.

[2] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:1124–1137, September 2004.

[3] Y. Boykov, O. Veksler, and R. Zabih. Efficient approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:1222–1239, November 2001.

[4] M. Cha and R. Phillips. Automatic track tracing in SAR CCD images using search cues. In *Signals, Systems and Computers (ASILOMAR)*, pages 1825–1829.

[5] M. Cha, R. Phillips, and M. Yee. Finding curves in SAR CCD images. In *ICASSP*, pages 2024–2027. IEEE, 2011.

[6] T. Cooke. Detection and classification of objects in synthetic aperture radar imagery. Technical Report DSTORR0305, Intelligence, Surveillance and Reconnaissance Division, Defence Science and Technology Organisation, Australia, 2006.

[7] C. V. Jakowatz, D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and P. A. Thompson. *Spotlight-mode synthetic aperture radar: a signal processing approach*. Kluwer Academic Publishers Norwell, MA, 1995.

[8] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:147–159, February 2004.

[9] S. Prince. *Computer Vision: Models Learning and Inference*. Cambridge University Press, 2012.

[10] T.-T. Quach, R. Malinas, and M. W. Koch. A model-based approach to finding tracks in SAR CCD images. In *IEEE Computer Vision and Pattern Recognition Workshops*, 2015.

[11] R. Steinbach, M. W. Koch, M. M. Moya, and J. Goold. Building detection in sar imagery. In *SPIE 9461, Radar Sensor Technology XIX; and Active and Passive Signatures VI, 94610E*, 2015.

[12] E. Turner, R. D. Phillips, C. Chiang, and M. Cha. Inserting simulated tracks into sar ccd imagery. In *Autumn Simulation Multi-Conference (AutumnSim'12)*. Society for Modeling & Simulation International, 2012.