TRACKING INFORMATION IN SAR IMAGE FORMATION AND CLASSIFICATION ALGORITHMS

Abhejit Rajagopal, Vincent Radzicki, Shivkumar Chandrasekaran, Hua Lee
Department of Electrical and Computer Engineering
University of California, Santa Barbra, USA 93106
{abhejit, vradzicki}@ece.ucsb.edu

ABSTRACT

Traditional target detection pipelines involve two sequential steps: the formation of a range-profile or likely-image, and the classification of likely targets within that image. Although it has been shown that target tracking in the RaDAR image-domain can be unnecessarily noisy, with more accurate and efficient implementations involving a direct analysis of the measured wavefield, image formation remains a desirable output in many applications due to its highly descriptive and interpretable nature. In this paper, we outline a mechanism for formalizing and accelerating this procedure in application-specific use cases. Enabled by recent advances in deep learning, we present a pipeline for automatically selecting an "optimal" filtered back-projection model, forming a likely-image, and performing target recognition and classification. The architecture allows practitioners to track and optimize the flow of information throughout the pipeline, enabling applications that utilize only intermediate outputs of the algorithm.

INTRODUCTION

The advent of radio detection and ranging (RaDAR) technology in the early 20th century was driven primarily by the need of allied strategic planners for an early aircraft warning system to provide them with important military intelligence [1]. The initial goal of early RaDAR systems was therefore detecting and tracking of both friendly and hostile aircraft. The resounding success that RaDAR systems found in this application generated widespread interest in the new technology and a search for other applications. Concurrently with these developments, research work on electromagnetic imaging techniques for scientific and medical instrumentation progressed rapidly [2]. The theoretical foundations between the two research areas were quickly recognized and work on developing radar imaging systems began soon afterwards [3, 4]. Radar imaging refers to estimating and forming spatial maps of the electromagnetic and other physical properties of objects and scenes probed by illuminating radar signals. The field of radar imaging covers a wide-variety of important engineering and scientific disciplines, and is utilized in a range of civil, medical, and military applications today.

As the capabilities of radar technology advanced from simple target detection and ranging to full-scale imaging, radar systems evolved to perform more complicated signal processing and analysis tasks. Tasks such as as automated pattern recognition and classification of radar data, for ex-

ample, became highly desirable in a number of applications. In remote sensing airborne Synthetic Aperture Radar (SAR) systems, scene classification can be particularly useful for determining the environmental characteristics and topography of a surveyed area, where the scenes could be classified as forest, desert, mountainous, urban, or rural [5]. In radar array imaging, automated target recognition (ATR) is a useful tool to detect specific targets, such as hostile aircraft, and classify the target into a set of pre-defined categories (i.e. aircraft model, size, etc). In high-resolution radar tomography, where objects are imaged from multiple positions over a virtual or physical array, classifying the material composition of objects and propagation media is an important problem for many civil engineering applications.

The methodologies used to perform these recognition and classification tasks has been an interesting development as well. Radar imaging systems will typically make measurements in the time or frequency domains, and the raw data will undergo significant processing to be mapped to the spatial domain for visual image display. Early techniques focused on interpreting the raw data for patterns to help classify features, such as the harmonic resonances of structures at specific wavelengths of electromagnetic illumination, and frequency dependent scattering parameters. Other techniques explored time-dependent features such as the micro-doppler effects of rotating or moving structures fixed to moving targets [6, 7]. In general, however, these early techniques have been limited to scenarios where targets are known and their scattering profiles can be explicitly modeled in the linear, Fourier domain [8, 9, 10]. While this is a reasonable "first-order" approach, these algorithms have historically failed to transition well to real-world and real-time scenarios where the number and distribution of targets cannot be guaranteed [11, 12].

As such, rather than processing the raw wavefield pattern, recent approaches have shifted their focus instead to processing and characterizing materials in the reconstructed image or inverse domain [13, 14]. In these typical SAR imaging systems, back-projection or filtered back-projection is used to reconstruct a scene from the measured complex wavefield pattern [10, 15]. With machine learning and computer vision algorithms, including deep neural networks, having demonstrated remarkable performance in detection and classification tasks when applied to optical imagery, it was expected that the application of these same algorithms to radar images would be straightforward [16]. However, many of these algorithms tend to focus on morphological and spatially coherent properties of objects that are not present in radar images, thereby limiting their success. Furthermore, the properties of radar images, which are intimately dependent on the image reconstruction algorithm, often exhibit imaging artifacts that can lead to misinterpretation by humans and algorithms alike. Still, the application of machine learning and classification algorithms to SAR imagery shows much promise in uncovering hidden patterns in data and measurements, which could significantly improve both classification performance and our ability to understand the information content of radar images quantitatively. Therefore, to enable high accuracy classification of targets in SAR imagery, it is essential to track and understand the flow of information throughout the image formation and classification pipeline.

In this paper, we discuss popular image formation algorithms, subsequent processing using neural networks, and detail an architecture which incorporates the two into a comprehensive learning framework. The presented network architecture and parametrized objective function enables practitioners to weigh the benefits of forming human-interpretable images against high-accuracy classification. In short, this work suggest that radar images that are appear semantically "focussed" to humans may not be the optimal representation for signal discrimination by a machine. By re-

moving this constraint on the image representation, we hope to achieve high accuracy ATR and target localization algorithms that can be tailored to user-specified applications and environments.

ADJOINT IMAGE FORMATION

In this section we briefly review some classical techniques for image formation using the adjoint operator. In particular, with appropriate assumptions, many electromagnetic imaging problems can be reduced to solving the follow linear system (Eq. 1):

$$d = Gm \tag{1}$$

where m captures the geometry and attenuation model of an environment, G represents the electromagnetic Green's function of the media, and d contains the known or measured data. In the case of active reflection-mode SAR, d contains both the known transmitted pulse or waveform and the received pulse (usually as a function of frequency), and m is the desired unknown that represents the reflection or absorption profile of the surveyed scene. From this linear-systems perspective, a reasonable estimate of m can be given by the least-squares solution (Eq. 2),

$$m = (G^H G)^{-1} G^H d \tag{2}$$

although in many cases $G^H G$ is invariably singular, ill-conditioned, and too large to easily invert. Classically, such conditions have encouraged practitioners to find suitable approximations of this equation. One common substitution is to eliminate the leading term by applying the approximation:

$$G^H G \approx I \tag{3}$$

which allows m to be computed easily as:

$$m \approx G^H d \tag{4}$$

where G^H is called the adjoint operator.

A. Back-Projection

Image formation traditionally involves using the measured data d to back-project the measured wavefield over either 1, 2, or 3 dimensions to construct a representative image of the scene m. The back-projection operator is typically selected as the inverse of the Green's function that defines forward-wave propagation in a homogeneous medium (i.e. backwards-wave propagation). For example, in Equation 4 the back-projection operator is chosen as the complex-conjugate of the Green's functions G and, in 3 spatial dimensions (x,y,z), takes the form:

$$G_i^H(r, \lambda) = \frac{1}{+j \cdot r \cdot \lambda} \cdot exp(+j \cdot 2\pi \cdot \frac{r}{\lambda}) \tag{5}$$

where r is the propagation distance of a coherent waveform, λ is the wavelength of light in an assumed medium, and j is the imaginary number $j^2 = -1$. This formulation permits a fast, simple, and vectorizable algorithm that can be tuned for real-time applications, as outlined below (Alg. 1).

It is worth noting that there are many ways to implement and speed up back-projection, with the presented pseudo-code being only one of many formulations that are optimized for parallelism and simplicity in understanding. In general, the complexity of back-projection stems from the fact that the Greens kernel changes for each "radial" point and operating frequency, with the pattern of how these functions are re-used being slightly more complicated. Typically, the explicit computation of these Green's functions is a bottleneck for the application of the adjoint operator, even if done in the frequency domain. As such, in practice the Greens kernel is often pre-computed with respect to a relevant antenna geometry, and re-used in all subsequent image formation tasks.

Algorithm 1 Image Formation using Back-Projection

procedure BP($grid$, ant, d) ▷ Given an evaluation grid, antenna array, and wavefield data
  $img \leftarrow$ zeros($grid.shape$)
  for $x \in grid.X$ **do**
    for $y \in grid.Y$ **do**
      for $z \in grid.Z$ **do**
        $pt \leftarrow (x, y, z)$ ▷ Point in scene
        $r \leftarrow ||ant.TX - pt||_2 + ||ant.RX - pt||_2$ ▷ Path-length between TX/RX
        $g \leftarrow \frac{1}{jr\bar{\lambda}} \cdot exp(2\pi jr\bar{\lambda}^{-1})$ ▷ Green's kernel for each $\bar{\lambda}_j$
        $img(x, y, z) \leftarrow sum(g(r, \bar{\lambda}) \cdot d)$ ▷ Apply adjoint operator
      end for
    end for
  end for
  return img
end procedure

B. Filtered Back-Projection

Although back-projection works well with limited modeling and computational effort, it is known that the resultant images are often unreliable for target identification. It is common, for example, for small scatters in particular geometries to be overly emphasized in the reconstructed image, particularly when imaging in the presence of multiple dielectric mediums. In particular, one of the issues with the back-projection model, besides its limitations in representing nonlinear systems, is that a single Greens kernel (i.e. derived for homogenous media) is applied throughout the reconstructed scene. For multi-layered heterogeneous media, this issue can greatly degrade the final image quality. One technique to overcome this is to modify the model to include environment-specific parameters, or spatial filters, that account for this inhomogeneity in a natural way. In this view, filtered back-projection can be viewed as a simple modification of the standard adjoint model to a system with the solution:

$$m \approx LG^H d \tag{6}$$

where the rows of L can be thought of as an application-specific environment or characterization model. Of course, the choice of L can be a difficult problem, but in general this allows for more degrees of freedom in the imaging model without sacrificing benefit of a fast algorithm. One such algorithm, which follows the structure of Algorithm 1 is presented on the next page in Algorithm 2.

Algorithm 2 Image Formation using Filtered Back-Projection

procedure FBP($grid$, ant, d, L) ▷ Given an evaluation grid, antenna array, wavefield data, and FBP model
  $img \leftarrow$ zeros($grid.shape$)
  for $x \in grid.X$ **do**
    for $y \in grid.Y$ **do**
      for $z \in grid.Z$ **do**
        $pt \leftarrow (x, y, z)$ ▷ Point in scene
        $r \leftarrow ||ant.TX - pt||_2 + ||ant.RX - pt||_2$ ▷ Path-length between TX/RX
        $g \leftarrow \frac{1}{jr\bar{\lambda}} \cdot exp(2\pi jr\bar{\lambda}^{-1})$ ▷ Green's kernel for each $\bar{\lambda}_j$
        $img(x, y, z) \leftarrow sum(L \cdot g(r, \bar{\lambda}) \cdot d)$ ▷ Apply modified adjoint operator
      end for
    end for
  end for
  return img
end procedure

C. FFT Technique

A main driver in the computational complexity of this back-projection algorithm is the complex operations involved in computing the Green's function. By discretizing the Green's kernel over propagation distance r and ignoring the amplitude attenuation component, a more efficient version of the algorithm can be computed. This possible modification involves altering the Green's kernel itself. In particular, practitioners have observed that the homogeneous Green's kernel (Eq. 5) closely resembles the commonly used Discrete Fourier Transform (DFT) basis functions:

$$f_i(\cdot) = exp(+j \cdot 2\pi \cdot k \cdot \frac{n}{N}) \tag{7}$$

where k represents the wave number, n is the sample index (e.g. in the time or spatial domain), and N is the total number of basis functions in the Fourier decomposition of the measured signal. Moreover, if N is restricted to be a power of 2, the fast Fourier Transform (FFT) can be utilized as a computationally efficient alternative formulation of back-projection, used to compute the reconstruction at each grid point. Furthermore, in this view the FFT formulation of SAR image reconstruction can be seen as an approximation of homogeneous Green's kernel by the Fourier basis, albeit one that enables a much faster computation, especially when the FFT is implmented directly in hardware.

PATTERN RECOGNITION

As discussed previously, it is both desirable and common for practitioners to localize, identify, and classify targets and materials using reconstructed images, rather than using the original wavefield data, even though mathematically these representations possess similar information. One reason for this might be that human operators can rely on geometric and spatial intuition to segment and identify objects when the information is represented spatially in a reconstructed image.

Specifically, the application of the previously described adjoint operator can be viewed as a low-pass filtering of the collected SAR wavefield data that spreads the measured information over the unambiguous range of a scene.

In this vein, various machine learning algorithms have been applied to pattern recognition tasks in SAR imagery, often incorporating intuition and priors from human operators that have been trained to interpret these images to produce application-specific analysis, such as military intelligence. While these tasks have historically utilized digital image-processing based denoising, segmentation, and classification algorithms [16], recently there has been significant interest in applying deep neural networks to these tasks [17]. Neural networks, which have emerged as a tractable approach to universal function approximation, with numerous applications in pattern recognition and image understanding, have been applied to a number of SAR datasets with varying levels of performance [18, 19]. One drawback, however, of current neural network architectures is that a majority of the existing machinery is built for real-valued inputs (e.g. intensities), rather than complex-valued inputs (e.g. wavefield data) [20, 21]. As such, a majority of neural network investigations utilize reconstructed magnitude images, which immediately reduces the amount of information that can be used to discriminate materials and targets in a scene.

Moreover, it should be noted that because the selected back-projection operator is not always invertible, the standard image-domain representation is not always optimal. As such, in addition to the various imaging artifacts that are introduced by simplifying assumptions in the system model, the reconstructed image itself is not always reliable for identification or characterization, perhaps due to the lack of information in this representation. That is, selected back-projection or filtered back-projection models that are semantically agreeable to human operators may not be the optimal intermediate representation of the wavefield data for characterization and classification tasks. By instead viewing the previously described image formation and pattern recognition algorithms (e.g. neural networks) as part of the same computation graph, it becomes clear that both the filter model and the neural network weights should be jointly optimized. An end-to-end classification and filtered-back projection algorithmic architecture is presented in Figure 1, on the following page.

In this architecture (Fig. 1), the image formation and pattern recognition task are optimized jointly using an objective function that values the interpretability of the image to both human operators and downstream classification networks. One such objective function could be parametrized as:

$$f_{image} = E(x, y) + \lambda V(y) \tag{8}$$

$$f_{detect} = ||z - z_{true}|| \tag{9}$$

$$f_{overall} = \alpha f_{image} + \beta f_{detect} \tag{10}$$

where x is either the output of standard back-projection or a known spatial map of the scene, y is the output of a learned filtered back-projection model, and z is the either the pixel-wise or collective classification result (e.g. the location and/or classification of a scatter in the scene).

The presented detection objective (Eq. 9) is fairly standard, and could be traded for a number of different objective (binary cross-entropy, mean-square error, log-liklihood, etc) depending on the application. Similarly, the image formation objective (Eq. 8) can match that of total-variation (TV) denoising, or other similar constraints commonly used in under-sampled image formation. For example, a popular choice is to pick (with an appropriate norm) a measure of anisotropic variance

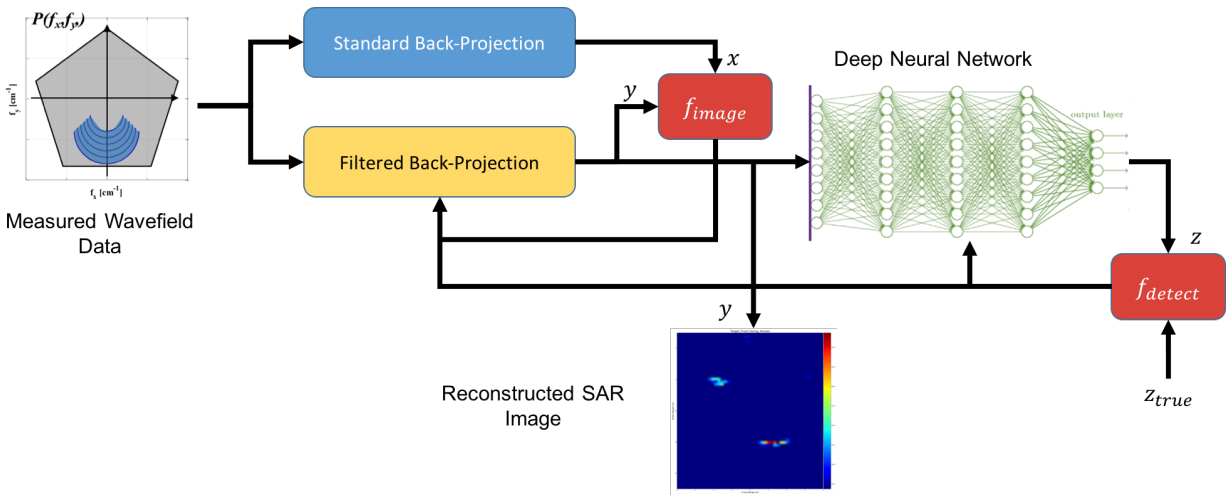

Figure 1: A deep neural network architecture that incorporates complex-valued image formation and complex-valued deep neural networks. The neural network at the end can be either a convolutional or fully-connected network that performs either pixel-wise or patch-wise classification and detections.

as:

$$E(x, y) = ||x - y|| \tag{11}$$

$$V(y) = \sum_{i,j} |y_{i+1,j} - y_{i,j}| + |y_{i,j+1} - y_{i,j}| \tag{12}$$

In this way, deep learning can be utilized to learn feature representations of the wavefield data that are semantically pleasing to human operators (e.g. via images), but still maintain a high level of discriminability to support their use in high-accuracy classification and target localization tasks. The degree to which this is achieved is then left to the user, who may select these parameters (α, β, λ) in real-time depending on their application requirements. For example, operators that require high-confidence classification may pick $\alpha = 0$ at the expense of losing visual interpretability in the image, while operators that require high-accuracy image formation may more carefully balance their selection of parameters.

CONCLUSIONS

We have reviewed the formulation of several popular image formation algorithms, and introduced an algorithmic architecture that suggests filter models for image formation should be tuned jointly with classification models. In particular, we present deep neural networks as an attractive candidate for the classification portion of the network, preceded by more-traditional filtered back-projection layers that can offer human-interpretable visual SAR imagery as a highly-descriptive intermediate output. Future work includes the incorporation of additional image formation and inverse-scattering models, the development of complex-valued neural network machinery, and the application of these to real-world terrain classification, material characterization, and target recognition tasks.

REFERENCES

[1] M. I. Skolnik, "Introduction to radar," *Radar Handbook*, vol. 2, 1962.

[2] S. Webb, *The physics of medical imaging*. CRC Press, 1988.

[3] A. Olte and S. Silver, "New results in backscattering from cones and spheroids," *IRE Transactions on Antennas and Propagation*, vol. 7, no. 5, pp. 61–67, 1959.

[4] D. L. Mensa, "High resolution radar cross-section imaging," *Boston, MA, Artech House, 1991*, vol. 1, 1991.

[5] L. Gomez-Chova, D. Tuia, G. Moser, and G. Camps-Valls, "Multimodal classification of remote sensing images: A review and future directions," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1560–1584, 2015.

[6] Y. Kim and H. Ling, "Human activity classification based on micro-doppler signatures using a support vector machine," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1328–1337, 2009.

[7] V. R. Radzicki, "Detection and imaging of micro-periodic motion with fmcw sensing systems," in *International Telemetering Conference Proceedings*, International Foundation for Telemetering, 2015.

[8] N. E. Huang, "Survey of remote sensing techniques for wave measurement," in *Measuring Ocean Waves: Proceedings of a Symposium and Workshop on Wave Measurement Technology; April 22-24, 1981; Washington*, p. 38, National Academies, 1982.

[9] E. A. Kasischke, D. R. Lyzenga, R. Shuchman, Y. Tseng, and B. S. Termaat, "The use of synthetic aperture radar to detect and chart submerged navigation hazards.," tech. rep., DTIC Document, 1982.

[10] J. W. Goodman, *Introduction to Fourier optics*. Roberts and Company Publishers, 2005.

[11] E. R. Stofan, D. L. Evans, C. Schmullius, B. Holt, J. J. Plaut, J. van Zyl, S. D. Wall, and J. Way, "Overview of results of spaceborne imaging radar-c, x-band synthetic aperture radar (sir-c/x-sar)," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 4, pp. 817–828, 1995.

[12] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *International journal of Remote sensing*, vol. 28, no. 5, pp. 823–870, 2007.

[13] T. R. Reed and J. H. Dubuf, "A review of recent texture segmentation and feature extraction techniques," *CVGIP: Image understanding*, vol. 57, no. 3, pp. 359–372, 1993.

[14] V. Dey, Y. Zhang, and M. Zhong, *A review on image segmentation techniques with remote sensing perspective*. na, 2010.

[15] H. Bischof, W. Schneider, and A. J. Pinz, "Multispectral classification of landsat-images using neural networks," *IEEE transactions on Geoscience and Remote Sensing*, vol. 30, no. 3, pp. 482–490, 1992.

[16] L. Capineri, P. Grande, and J. Temple, "Advanced image-processing technique for real-time interpretation of ground-penetrating radar images," *International Journal of Imaging Systems and Technology*, vol. 9, no. 1, pp. 51–59, 1998.

[17] M. Wilmanski, C. Kreucher, and J. Lauer, "Modern approaches in deep learning for sar atr," in *SPIE Defense+ Security*, pp. 98430N–98430N, International Society for Optics and Photonics, 2016.

[18] K.-S. Chen, W. Huang, D. Tsay, and F. Amar, "Classification of multifrequency polarimetric sar imagery using a dynamic learning neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 3, pp. 814–820, 1996.

[19] N. Kussul, M. Lavreniuk, A. Shelestov, and B. Yailymov, "Along the season crop classification in ukraine based on time series of optical and sar images using ensemble of neural network classifiers," in *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*, pp. 7145–7148, IEEE, 2016.

[20] D. A. Morgan, "Deep convolutional neural networks for atr from sar imagery," *Proceedings of the Algorithms for Synthetic Aperture Radar Imagery XXII, Baltimore, MD, USA*, vol. 23, p. 94750F, 2015.

[21] M. Wilmanski, C. Kreucher, and A. Hero, "Complex input convolutional neural networks for wide angle sar atr," in *Signal and Information Processing (GlobalSIP), 2016 IEEE Global Conference on*, pp. 1037–1041, IEEE, 2016.