Efficient Generation of Image Chips for Training Deep Learning Algorithms

Sanghui Han[a], Alex Fafard[a], John Kerekes[a], Emmett Ientilucci[a], Michael Gartley[a], Andreas Savakis[b], Charles Law[c], Jason Parhan[d], Matt Turek[c], Keith Fieldhouse[c], and Todd Rovito[e]

[a]Rochester Institute of Technology, Center for Imaging Science, Rochester, NY
[b]Rochester Institute of Technology, Department of Computer Engineering, Rochester, NY
[c]Kitware Incorporated, Clifton Park, NY
[d]Rensselaer Polytechnic Institute, Troy, NY
[e]Air Force Research Laboratory, Wright-Patterson Air Force Base, OH

ABSTRACT

Training deep convolutional networks for satellite or aerial image analysis often requires a large amount of training data. For a more robust algorithm, training data need to have variations not only in the background and target, but also radiometric variations in the image such as shadowing, illumination changes, atmospheric conditions, and imaging platforms with different collection geometry. Data augmentation is a commonly used approach to generating additional training data. However, this approach is often insufficient in accounting for real world changes in lighting, location or viewpoint outside of the collection geometry. Alternatively, image simulation can be an efficient way to augment training data that incorporates all these variations, such as changing backgrounds, that may be encountered in real data. The Digital Imaging and Remote Sensing Image Image Generation (DIRSIG) model is a tool that produces synthetic imagery using a suite of physics-based radiation propagation modules. DIRSIG can simulate images taken from different sensors with variation in collection geometry, spectral response, solar elevation and angle, atmospheric models, target, and background. For our research, we selected ground vehicles as target objects and incorporated the Simulation of Urban Mobility (SUMO) model into DIRSIG to generate scenes with vehicle movement. SUMO is a multi-modal traffic simulation tool that explicitly models vehicles that move through a given road network. Using the combination of DIRSIG and SUMO, we can quickly generate hundreds of image chips, with the target at the center with different backgrounds. The simulations generated chips with vehicles and helicopters as targets, and corresponding images without targets. Using parallel computing, 120,000 training images were generated in about an hour. Some preliminary results show an improvement in the deep learning algorithm when real image training data are augmented with the simulated images.

Keywords: Remote Sensing, Simulation, Convolutional Neural Network, Target Detection, Deep Learning, Data Augmentation, Synthetic Imagery

1. INTRODUCTION

Large amounts of training data are required to develop robust deep learning algorithms for target detection using satellite imagery. Obtaining the amount of data required that is also labeled with known truth information along with labeled imaging conditions can be challenging. Realistic image simulation provides a cost effective method to overcome this challenge. In this research, we generated millions of small images or "image chips" that contained a defined number of pixels for a target to use or modify for the purpose of training a convolutional neural network. These image chips could be generated quickly, for various collection times, atmospheric conditions, and sensor specifications as only the radiance values for the required number of pixels were calculated. Since the synthetic images have perfect truth and are labeled with known imaging conditions, certain aspects of the imaging process can be isolated to find parameters that most affect the deep learning algorithm. For instance, the atmospheric effects can be identified as a factor that the deep learning algorithm is sensitive to if results change from including or excluding images chips from a specific atmospheric model used in the simulation. This can give information on where to focus efforts in improvement or modification of the algorithm.

2. APPROACH

We used several models to generate the synthetic images and changing backgrounds. The synthetic image generation tool we used was the Digital Imaging and Remote Sensing Image Generation (DIRSIG)[1] model that was developed by the Digital Imaging and Remote Sensing (DIRS) Laboratory at Rochester Institute of Technology (RIT) over the last 30 years. DIRSIG can work with plugins from other models, and we used the Simulation of Urban Mobility (SUMO),[2] to quickly change the vehicle and helicopter location and surrounding background geometry in each image chip in order to incorporate local environment effects of buildings, vegetation and other scene geometry on targets.

2.1 Digital Imaging and Remote Sensing Image Generation

The image chips were generated using DIRSIG, which uses physics-based principles and reverse ray tracing to calculate a radiometric value at a pixel given the input parameters. Figure 1 shows the different parameters that are incorporated in the final synthetic image.

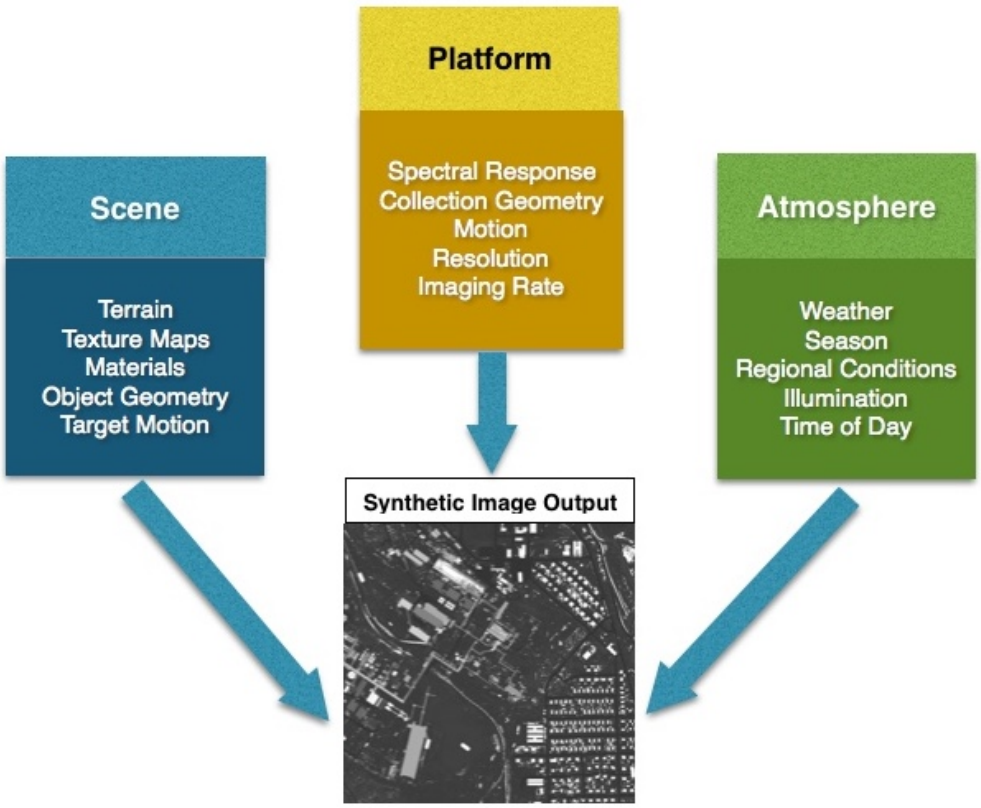

Figure 1. DIRSIG input components.

We simulated sensors that generate passive broadband images and detect signals in the visible spectrum. DIRSIG incorporates atmospheric effects in the visible and near-infrared spectrum, using an external module called Moderate Resolution Atmospheric Transmission (MODTRAN).[3]

MODTRAN is a computer code that predicts optical measurements through the atmosphere. It computes line-of-sight spectral transmittance and radiance using radiation transport physics assuming horizontal homogeneity through each constituent vertical profiles of the atmosphere. It can model climatology data and solves

the radiative transfer equation taking into account molecular and particulate absorption/emission and scattering, surface reflections and emission, solar/lunar illumination, and spherical refraction. In DIRSIG, the radiometry engine accounts for the multiple scattering of energy from the sun using the atmospheric database file that is generated using an executable that interfaces with MODTRAN, and takes into account scene geometry, material files, and climatology information.[4]

The scene is described using terrain information, texture maps, object files, and geometry list files. The object files are 3-D polygon models that can be created using design tools such as AutoCAD or Blender3D. Each object is associated with a list of materials that reference optical properties of each material that are used to drive the radiometric prediction. The geometry list files have the x, y, and z coordinates of the objects to place them within the scene. The texture maps are files that have the background look and feel of the scene on which the 3-D object geometries can be placed, but reference a spectral database for its radiometric information based on the materials assigned to the texture. It then takes in the terrain information to give variation in the ground height at each pixel location.

The platform and its imaging components are described by the spectral response function of the detectors, the motion of the platform, height and focal length of the imaging system above the scene, and the description of the detector array. The information contained in the platform files determine the ground sampling distance (GSD) of the output image. The spectral response function of the detectors can be defined and specified for panchromatic or multispectral images.

The scene, platform, and atmosphere information that is defined by the user triggers a set of radiometry solvers that accounts for multi-scattering of energy from the sun, moon and sky, and calculates a radiance value for a scene at each pixel. The amount of time it takes to generate a single image is dependent on the number of pixels in the image and complexity of the geometry.

In this research we used the imaging parameters of Worldview 1 and Worldview 2 satellite sensors, but simulated a tracking mount attachment in DIRSIG. The tracking mount follows a target object and and generates multiple images at a defined rate over a time interval, in order to simulate local environment effects of buildings and other shadowing effects. The output of a tracking mount simulation if an image with the specified target is placed at the center. Even if there is obscuration due to trees or buildings, we know it is a positive image chip. We used an imaging rate of 0.5 Hz over a 100 second interval to generate 50 chips per mount. We simulated ten mounts placed on a platform to generate a total of 500 chips per run.

2.2 Scene and Objects

Two different scenes were used to provide the background material, terrain, and geometry for variations from one image chip to another. One scene was of Irondequoit, NY, which is a suburban area with mostly residential buildings, abundant trees and vegetation, and typically suburban transportation infrastructure such as roads and sidewalks.[5] Another scene used was of Trona, CA which is an industrial area in the desert. It has large open areas of sand and soil along with industrial plants and large single-story buildings. It also has railway lines and unpaved roads. These two scenes were used with the tracking mount platform to create image chips that had the target in the center, but surrounded with different background and geometry for each chip in a single simulation instance.

The files for the scenes were developed by the DIRS Laboratory and contained geometry files for different types of vehicles and road networks. The pre-exisiting files were leveraged to generate the realistic background for the training data using the SUMO plug-in. Object files of Russian military helicopter models were created in Blender3D as obtaining sufficient real image training data for this target was a great challenge. The vehicles were placed on the road networks of both scenes, since both were areas where vehicles are typically present. The helicopter targets were placed only in the Trona, CA scene as Russian military helicopters are not usually found in suburban areas.

2.3 Simulation of Urban Mobility

The movement files used for the tracking mount simulation provided the location information of the target object at each time interval in order to generate the image chips with the target centered at that point. These files were generated by SUMO, which is a traffic simulation tool that incorporates a given road network and vehicles as input and models the movement of the vehicles within the road network. This placed the vehicles at realistic and likely locations.

Figure 2 shows the Trona, CA road network that was used as input into SUMO to generate some of the movement files. The movement files contain information about a targets initial position, then a list of horizontal, vertical, height, and rotational displacements from the initial position for each increment of time. For the helicopter targets, the same movement files were used to generate the changing background, but with some modification to allow helicopters to be in open areas. The initial elevation information was also varied to simulate both grounded and airborne helicopters. In order to simulate rotating rotors, the rotors and the fuselage were treated as separate objects. The rotors were given random rotations to ensure there were many variations of rotor placements in relation to the fuselage.

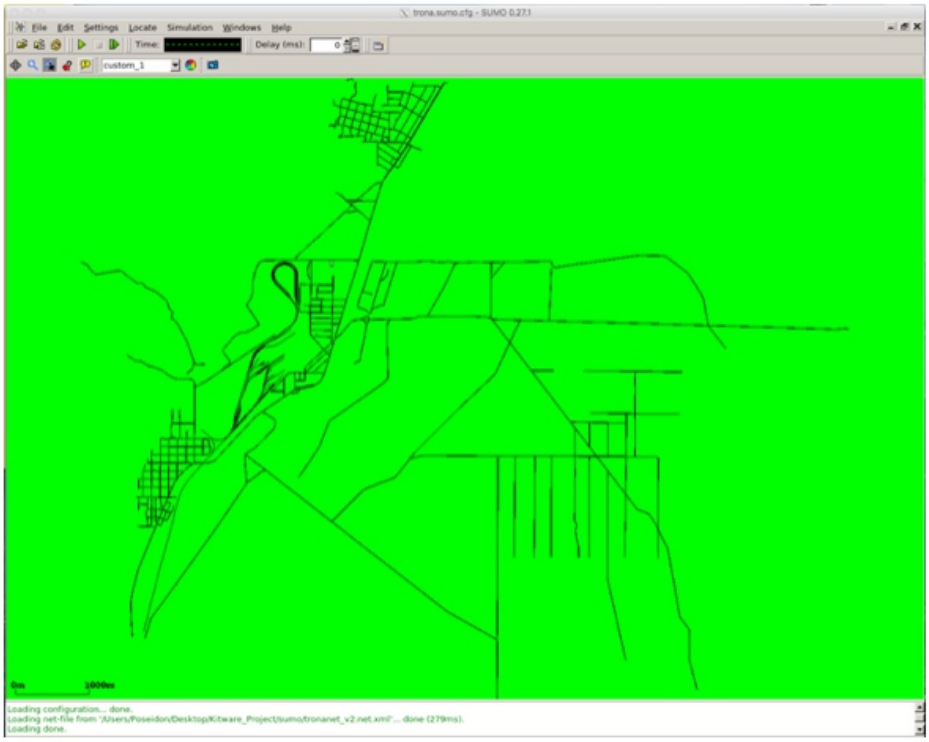

Figure 2. Road network of Trona, CA used as input for SUMO to generate movement files to use in DIRSIG.

2.4 Parameter Variation

There are a myriad of parameters that affect an image. The parameters we chose were selected as the variables most likely to change the output of deep learning algorithms. The background and geometry were varied within a single simulation instance. The platform, atmosphere, and solar elevation/angles were varied from one instance to another. In essence, each simulation contained changing background and object geometry, and was replicated to vary atmosphere, solar elevation/angle and the type of sensor. The type of sensor that was simulated determined whether a panchromatic or RGB image was rendered. For example, Worldview 1 only has panchromatic capabilities, and so only panchromatic images were generated. Worldview 2 on the other hand had panchromatic

and multispectral capabilities and so both panchromatic and RGB images were generated. Table 1 shows the parameters that were taken into consideration and the variations produced in the image chips.

Table 1. Parameter Variations

Parameters	**Variations**
Sensor	Worldview 1, Worldview 2
Atmospheric Model	Mid-latitude winter and summer, Sub-arctic winter and summer, Tropical
Solar Elevation/Angle	30 minute increments in 24-hour time period
Spectral	RGB, Panchromatic
Scene	Suburban, Desert, Industrial
Background	Road, Grass, Asphalt, Sand, Soil
Geometry	Residential buildings, Trees, Industrial infrastructure, Vehicles, Helicopters

For the Worldview 2 platform, panchromatic and RGB images with a 0.46m ground sampled distance (GSD) were rendered. This simulated the final products available from DigitalGlobe, instead of pre-processed images. This was because RGB images that are not pan-sharpened have a 2m GSD, and vehicles would not be distinguishable and unusable as a target at this resolution. Other commercial imaging satellite platform configurations were considered, but the DigitalGlobe Worldview 1,2 were the only platforms that produced images with the resolution that allowed both vehicles and helicopters as viable targets. We plan to address other commercial sensing platforms in the future using larger targets such as ships.

When using variations in the atmospheric models, the date of collection for the winter models was set to 10 January 2010, and for the summer and Tropical models the date was set to 10 July 2010. The collection times were set to the entire 24-hours, although the real images collected at night would be all black and consist only of noise. This was in order to preserve the images simulating the collection times with the longest shadows. The season, geographical location, and atmospheric model will have an effect on the shadow effects due to solar elevation change. For example, mid-latitude summers in Trona, CA has a sunrise time between 5-6am. The chips produced during this collection time period provided a lighting variation in the training data were of great interest. Therefore, the determination of image chips with unrealistic radiance values was done post-processing, and were not used.

3. RESULTS

A total of 5.7 million chips with and without targets were generated for the parameters listed in Table 1. The vehicle chips generated included variations in vehicle color and model. The helicopter chips included five different Russian helicopter models and both targets had varying background, weather and surrounding environment effects such as shadows and partial occlusion present from buildings and trees. Figures 3 and 4 show examples of some of the different chips that were generated at various times under mid-latitude summer conditions simulating a Worldview 2 sensor.

3.1 Generation Speed

Each simulation instance that generated 500 image chips took less than 15 minutes to run on a single CPU. There were 240 simulation instances that were created for a single simulation scene and sensor that consisted of 48 collection times and 5 atmospheric conditions. They were run simultaneously on the RIT research cluster as well as the Air Force Research Laboratory High Performance Computing facility machines. On a typical day with the priority for computational resources given to students, it took about an hour for all 240 instances to run on the

Figure 3. Examples of RGB image chips generated using Irondequoit, NY as background scene with variations in vehicle color and type.

RIT research cluster. Each image chip had an corresponding negative image chips that did not have the target present. These chips had identical backgrounds as the positive image chips containing the target object. They were slightly larger than the positive image chips, for example The larger negative patch size allowed multiple positive chips to be extracted from a single negative example.

We were able to incorporate diverse variations in atmospheric conditions, collection time, background and surrounding geometry where one simulation could generate 120,000 chips in about an hour. Using the existing scenes for Trona, CA and Irondequoit, NY as background, we can create different simulations with more variations in target models in the future relatively easily.

3.2 Limitations

It is impossible to simulate all the complex real world phenomena such as atmospheric effects and noise that affects an image. Furthermore, we have incomplete knowledge of sensor characteristics and unknown post-processing of images from a real sensor platforms. These factors that are not captured in the simulation can changes the final output of the synthetic image, which can affect the training results of a deep learning algorithm. Approximations of the real image can be improved by adding noise or other imaging artifacts, but there will always be differences that can not be duplicated.

3.3 Impact on Deep Learning Object Detectors

We have seen promising improvements in deep learning detector performance after incorporating additional synthetic training data. We performed one preliminary experiment to develop a detector for a subclass of helicopters with five blades, as illustrated in Figure 5. In part, this experiment was constructed to understand the value of

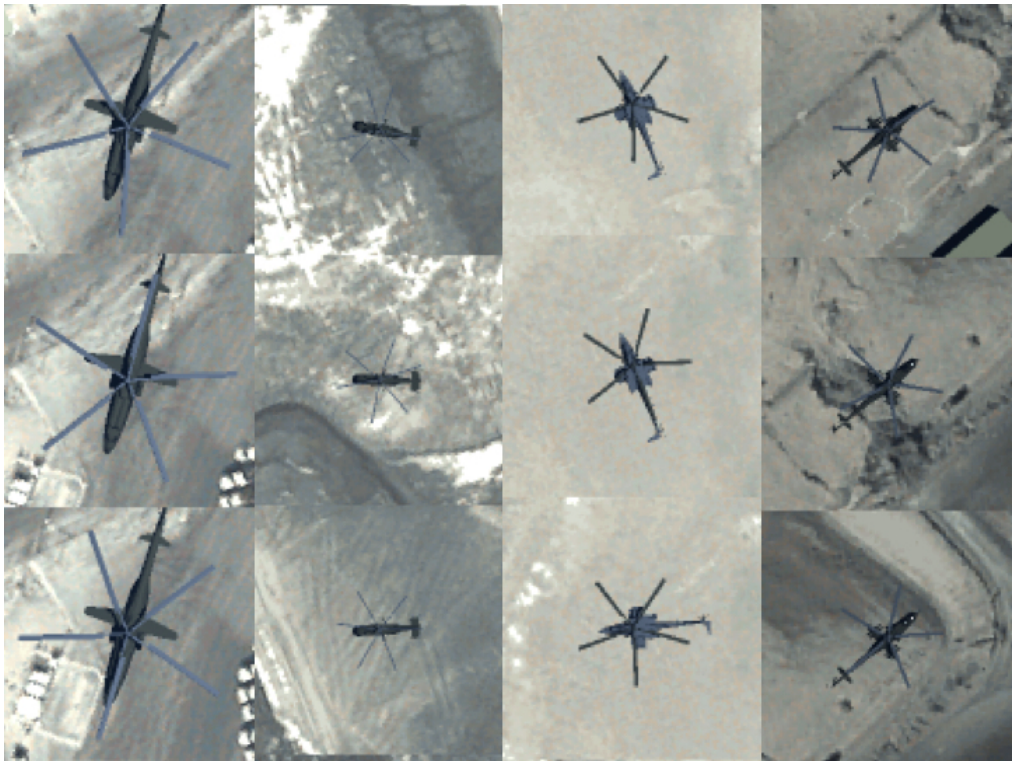

Figure 4. Examples of Panchromatic image chips using Trona, CA as background scene with variations in rotary-wing aircraft models simulated at airborne height of 3600ft.

synthetic training data for relatively fine-grained classification tasks, where the collection of sufficient training data to build a deep learning detector would be particularly challenging. We used 3829 real helicopter examples as training data with 23218 real negative examples randomly sampled from the background. 11665 synthetic helicopter examples were generated with DIRSIG along with a similar number of negative synthetic examples.

Figure 5. Synthetic five-bladed helicopters in flight produced using DIRSIG.

The addition of synthetic training data increased probability of detection from approximately 30% to 60% false alarm rate of 4 false positives per square km. Figure 6 contains receiver operating characteristic (ROC) curves from this experiment. Figure 7 illustrates the output of the detector trained with and without additional synthetic data on an example DigitalGlobe image.

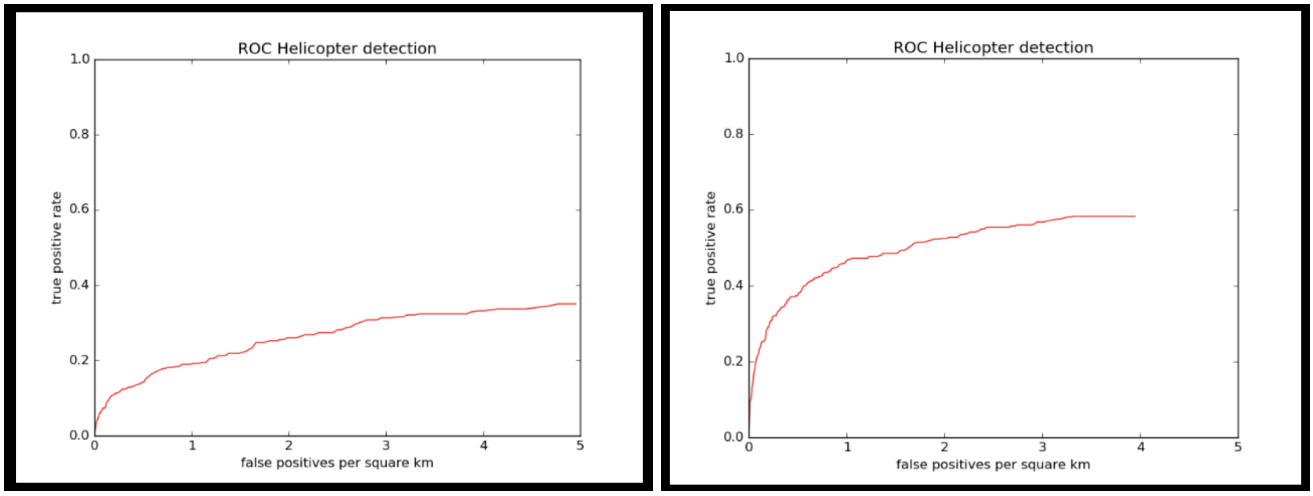

Figure 6. ROC curve for five-bladed helicopter detection before additional DIRSIG training data (left) and after additional DIRSIG training data (right).

Figure 7. Example output of detector algorithm trained without additional DIRSIG data (left) and trained with additional DIRSIG data (right).

4. FUTURE DIRECTIONS

In the future, we can simulate other commercial satellite imaging systems such as Planet Labs[6] or Terra Bella[7] for appropriate targets and scenes. These sensors were not considered for vehicle or helicopter targets because the images did not have the appropriate resolution. For example, a Planet Labs image has a GSD of 5-6m, and a vehicle is not distinguishable at this resolution. However, larger objects such as ships or airplanes, can be detected, and our next step can be to use them as targets. The scenes that were used to provide background

material and geometry were for suburban and industrial areas. Another scene under consideration is Tacoma, WA which is a coastal harbor area that provides a realistic background to place ships as targets.

ACKNOWLEDGMENTS

We would like to thank the Air Force Research Laboratory High Performance Computing facility for their efforts and provision of computational resources.

REFERENCES

[1] DIRS Laboratory, "DIRSIG Reference Docs." http://www.dirsig.org/docs/new/intro.html. (Late Updated: 21 July 2016).

[2] Hilbrich, R., "DLR - Institute of Transportation." http://www.dlr.de/ts/en/desktopdefault.aspx/tabid-9883/16931_read-41000/. (Last Accessed: 31 October 2016).

[3] Spectral Science Incorporated, "MODTRAN." http://modtran.spectral.com/. (Last Accessed: 31 October 2016).

[4] J.R. Schott, S.D. Brown, R. R. H. G. G. R., "An advanced synthetic image generation model and its application to multi/hyperspectral algorithm development," *Canadian Journal of Remote Sensing* **25**, 99–111 (1999).

[5] Emmett J. Ientilucci, S. D. B., "Advances in wide area hyperspectral image simulation," in [*Targets and Backgrounds IX: Characterization and Representation*], Wedell R. Watkins, Dieter Clement, W. R. R., ed., *Proc. SPIE* **5075**, 110–121 (2003).

[6] Planet Labs Incorporated, "The Planet Platform." https://www.planet.com/products/platform/. (Last Accessed: 29 November 2016).

[7] Terra Bella Google Company, "Terra Bella Satellites." https://terrabella.google.com/?s=about-us&c=about-satellites. (Last Accessed: 29 November 2016).