2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN) | 978-1-6654-9624-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/IPSN54338.2022.00025

SelfieStick: Towards Earth Imaging from a Low-Cost Ground Module Using LEO Satellites

Vaibhav Singh
Carnegie Mellon University
vaibhav3@andrew.cmu.edu

Osman Yağan
Carnegie Mellon University
oyagan@andrew.cmu.edu

Swarun Kumar
Carnegie Mellon University
swarun@cmu.edu

ABSTRACT

Real-time access to overhead Low-Earth Orbit (LEO) satellite imagery from a handheld device can have transformative applications: tracking wild-fire, natural disasters and weather events. Today, real-time access to images from LEO satellites overhead is challenging to obtain. LEO satellite ground receivers are bulky, expensive and sparsely deployed in the world. Despite the exponential increase in LEO small satellites orbiting the planet today – there is a significant time gap between an image capture on such a satellite and users who need it the most in remote and ecologically-sensitive regions.

This paper presents SelfieStick, a novel satellite receiver system that explores reducing this barrier of access to real-time satellite imagery data using a single low cost (< $ 30) tiny receiver. SelfieStick's core approach takes advantage of the multiplicity of overhead Low-Earth Orbit satellites due to their exponential rise in recent years. While signals from such satellites may be individually weak, especially at a low-cost receiver, SelfieStick stitches together noisy RF captures containing underlying images of the same part of the Earth across many such satellites to generate clean Earth images. This is made possible by combining weak signals in the RF domain (rather than the traditional image domain) after appropriately transforming and aligning the RF signals accounting for different satellite perspectives, their orbits and wireless channels. A detailed experimental evaluation on the RTL-SDR platform on satellite captures from the NOAA constellation demonstrates a PSNR improvement of 5 dB through combining of images across 10 satellites.

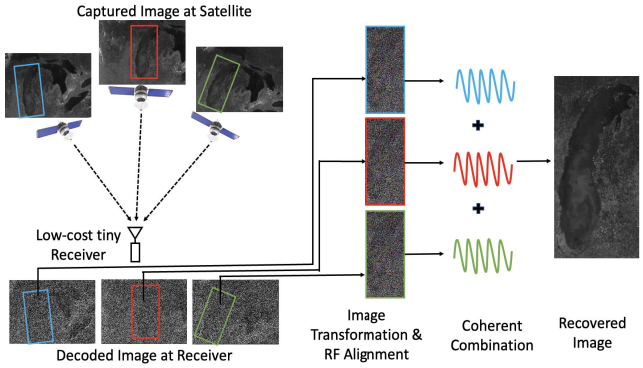

Figure 1: *Approach*: SelfieStick captures RF signals containing images of the same region of Earth from many satellites at a low-cost receiver. While these signals are individually too noisy to decode, SelfieStick transforms and combines them to obtain a cleaner image.

1 INTRODUCTION

"Our selfie stick, if you will, is 150 million miles long"
– Elsa Jensen (on the NASA Perseverance rover)

This paper explores building a low-cost (under $30) ground module that listens to overhead Low-Earth Orbit (LEO) satellites to build a real-time image of the Earth. Today, there are over 200 Earth observation satellites in orbit [7] equipped with varied sensors that measure the properties of the Earth and its atmosphere on a daily basis. This number is growing exponentially and very soon, nearly every point on the Earth's surface may have tens of LEO satellites passing overhead [11] at any given time. Yet ground stations for such LEO satellites has simply not kept up with this pace [46] – they are expensive, bulky and relatively sparse – very few ground receivers exist in remote regions or the developing world. In other words, collecting a real-time image of the Earth from overhead LEO satellites, especially in remote or ecologically-sensitive regions – whether for tracking real-time weather events, natural disasters, or simply for research – is a challenging proposition.

While private players like Planet Labs, DigitalGlobe and IKONOS [1, 4] operate a constellation of LEO satellites, providing fine grained earth images on demand, they do so at significant cost. A more hands-on approach is for users to either rent and schedule ground stations for reception[2, 6] or deploy their own ground station infrastructure [49, 50] and decode the receptions from these satellites. However, this approach too requires significant investment from the users in terms of cost and/or space for deployment. More recent efforts have strived to bring down the cost of ground infrastructure [40, 44, 46], yet still requires multiple distributed receivers.

This paper presents SelfieStick, a system that explores the feasibility of designing a single tiny receiver which costs a few tens of dollars and is small enough to be deployed anywhere (including indoors) to obtain recently captured satellite images sent by overhead LEO satellites. Naturally, receptions from individual satellites at a low-cost receiver are likely extremely weak – nearly undecodable. SelfieStick's key idea exploits the ever-increasing multiplicity of overhead LEO satellites. It uses weak receptions across many satellites, each containing images of the same region of the Earth taken from many perspectives to generate clean(er) images of that region (see Fig. 1). We implement and evaluate SelfieStick on extensive LEO satellite traces collected from an RTL-SDR receiver listening to the NOAA constellation. Our results reveal a PSNR improvement of 5 dB through combining signal captures from 10 satellites.

A naïve approach to build SelfieStick would simply map each satellite's signal to corresponding (potentially noisy) images. Of course, these Earth images are taken from different perspectives (often called "multi-view") due to different satellite orbits and orientations. One can then use the rich literature on multi-view image fusion [22, 53] to combine these images across diverse perspectives, despite some noise (e.g. with advanced image denoising [43]).

978-1-6654-9624-7/22/$31.00 ©2022 IEEE
DOI 10.1109/IPSN54338.2022.00025

Unfortunately, this approach fails in SelfieStick's context due to an important challenge – extremely high noise. To see why, note that most satellite modulations encode per-pixel information in the phase of received signals [30, 32, 39]. Such modulations face fundamental decoding limits from phase alone when signal power is well below noise power [8] – an exceedingly common situation for low-cost radios receiving satellite signals. In other words, if SelfieStick attempts to map highly noisy RF signal receptions to individual images – these images would simply look like uniformly-at-random noise, with negligible useful information or features.

To address this challenge, SelfieStick does not even attempt to decode images captured by individual satellites from their RF receptions in the first place. Instead, it makes corrections to images *indirectly* through processing the received RF captures, transforming and correcting the RF captures for underlying image distortions, differences in perspectives due satellite orbit, etc. In other words, SelfieStick proceeds in three key steps, once potentially noisy RF samples from different LEO satellites have been gathered: (1) *Satellite Perspective Estimation:* First, it estimates differences in perspectives between the diverse points-of-view from which each satellite captures the region of interest on the Earth. (2) *Image Alignment in RF:* Second, it processes the RF captures to indirectly eliminate these perspective differences in the underlying images. (3) *Coherent Multi-View Combination:* Finally, SelfieStick coherently combines the transformed RF signals to obtain a cleaner RF capture that map to a cleaner image. We expand on these steps below:

Satellite Perspective Estimation: In an ideal world where all overhead satellites capture an identical image of the Earth, one could perhaps simply hope to add up the corresponding RF samples from these satellites coherently. Unfortunately, satellites view the same region of Earth from different perspectives. The traditional computer vision approach to deal with this problem (e.g. in multi-view fusion [22, 53]) is to look for feature points in the underlying images to reverse engineer the perspective transform between them. As stated previously though, SelfieStick does not simply have direct access to the individual satellite images due to noise.

An alternate approach that SelfieStick takes is to use the known location of satellites to infer these transforms. Specifically, for most LEO satellites, their orbital parameters (i.e. 3-D location of the satellite over Earth) at any point in time is known to a sufficiently high degree of precision. Yet, one key attribute is often missing and unknown to users on the ground – the orientation of the satellite, given that satellites often experience location dependent attitude control as well as tumbling (in some cases) during its orbit [25, 36]. To address this challenge, SelfieStick leverages the fact that these orientation changes due to satellite location/tumble are reasonably predictable over time-scales of several days, mostly since satellites experience very little air resistance. SelfieStick can therefore make reasonable estimates of satellite orientation from its previous values at different times, estimated, say from clean captures obtained at an earlier time and/or location from a high-quality RF ground station. SelfieStick achieves this by tracking prominent landmarks found in most regions on Earth across multiple images. Sec. 4 describes our data-driven approach that estimates the appropriate image perspective transforms using previously captured satellite images and orbital parameters.

Image Alignment in the RF Domain: Having computed the image transforms, SelfieStick must apply these indirectly on the RF receptions. Doing so requires a satellite modulation-specific approach, since each modulation maps captured pixels into different RF samples. Given that SelfieStick is primarily evaluated on NOAA satellites (mainly due to public data availability), we detail this approach for the popular Automatic Picture Transmission (APT) modulation (we describe extensions to other modulations in Sec. 9).

At a high level, our approach must deal with many challenges in how pixels are mapped to samples. First, we must uncover the RF samples along two different captures that correspond to the same pixel. This becomes challenging for modulations such as APT where the same pixel is often "smeared" across multiple RF samples. Second, given that most modulations encode data in phase, SelfieStick must apply image transforms that are often matrix multiplications of pixels indirectly on RF samples to appropriately manipulate their phase in a consistent manner. Intuitively, SelfieStick achieves this by relying on useful properties of RF signal captures that are essentially complex numbers – for example multiplication of two RF samples is akin to the addition of their phases. Sec. 5 describes how each of these challenges are addressed to transform RF captures so that their underlying images are aligned to a common perspective.

Coherent Multi-View Combination: Next, we need to coherently combine the perspective-corrected RF captures to obtain a clean image. Prior to doing so, SelfieStick must estimate and account for time, frequency and phase offsets that result from the fact that each signal capture was obtained at different times from different satellites at the receiver as well as wireless channels and Doppler shifts. Sec. 6 describes how we address these challenges.

Limitations: A few key limitations of our system are: (1) Our system relies on multiple overhead satellites. At present, the scale of global satellite deployments allow for only 0-3 at a given time and 5-6 over a 3 hour period [23]. We expect this limitation to be temporary though as LEO satellite deployments rapidly scale [11]. (2) We assume the modulation of the satellites and their orbital parameters are known and satellite image data is not encrypted, as is the case for many public satellites. Our system's performance can be modulation-specific though and we address this in Sec. 9.

Evaluation: We implement and evaluate SelfieStick on an RTL-SDR (costing 25 USD) that can be connected to a Raspberry PI or the user's laptop. These are connected to EXS136SMI Laird technologies antennas [15] allowing for a form factor of 9.95 by 1.20 centimeters and not requiring specialized installation. We combine signal captures from up to 10 NOAA LEO satellites. Our results reveal the following improvements relative to a one-satellite baseline:

- An average improvement in received SNR of 8.4 dB.
- An increase in average SSIM from 0.53 with one satellite to 0.83 with 10 satellites
- An average improvement in PSNR of 5 dB.

Contributions: This paper contributes: (1) A novel system to obtain real-time overhead LEO satellite Earth-images from a low-cost ground receiver; (2) An approach to coherently combine individually weak satellite receptions in the RF domain, despite containing images taken from different perspectives; (3) A detailed system evaluation demonstrating significant image quality improvements.

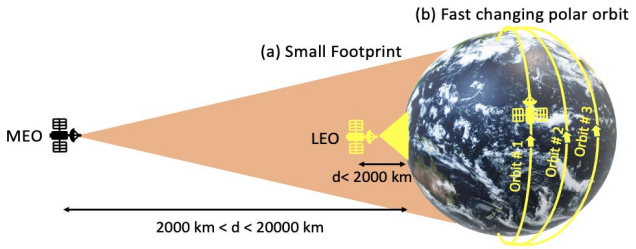

Figure 2: LEO satellites (a) operate in low altitudes and hence offer only a small angular view of the Earth, and (b) operate in non-overlapping sequential polar orbits and complete one rotation in just 100 min.

2 PRIMER

In this section, we describe some of the characteristics of LEO satellites: what makes them suitable for Earth observation and why their ground stations are expensive.

LEOs are fast and fleeting: LEO satellites orbit the earth between 160-2000 km in altitude, with most Earth observation satellites following a sun-synchronous polar orbit (Fig.2). The proximity of the satellites to the earth surface results in a small footprint subtended by them at any time. Due to their low altitude, small footprint and high speed, LEO satellites are visible at any point on earth only for a small duration of time. They complete one orbit in about 60-120 minutes depending on their altitude, and do not frequent the same geographic location on earth more than ~ 4-5 times a day (for non-polar regions).

LEO ground stations are expensive: Although LEO satellites are closer to the Earth than MEO satellites, the lower path loss they enjoy relative to these satellites are somewhat offset by the fact that most LEO satellites are small or cubesats. The small size of these satellites imposes strict constraints on the power that can be emitted from the their transmitter. Thus, to receive signals from these satellites, one must deploy an expensive ground station with an unobstructed view of the sky, comprising of a large antenna and a bulky rotator setup to keep track of the satellite's trajectory.

LEOs are growing in popularity: While Geo-synchronous (GEO) orbits were traditionally preferred for weather monitoring, there has been a progressive transition to low earth orbits for Earth observation [4, 7]. This trend can be seen across the board with government Earth observation constellations like NOAA, LANDSAT, SENTINEL as well commercially owned constellations like WorldView, Flock, SkySat. There are three factors driving this transition: (1) Proximity to the earth allows fine grained imagery with advertised resolutions as good as 4 m^2 per pixel; (2) LEO satellites allow fine grained coverage of different regions of the earth in single orbit; and (3) Applications like Synthetic Aperture Radar (SAR) require relative motion between the satellite and the imaging scene.

Latency from LEO imaging to down-linking: The transition to low earth orbits also necessitates ground stations to be extensively available. Most LEO satellites store data on board and downlink data when within communication range of a specific ground station (mostly deployed around the poles). This entails a high latency (~ 90 minutes) between capture and down-linking [46]. Deploying more base stations to reduce this latency is an option, but comes at increasing cost. Thus, there is a need for cheap and easily deployable ground station infrastructure to take full advantage of the potential that these Earth observation satellites offer.

3 SELFIESTICK DESIGN AND CHALLENGES

In this section, we motivate our choice of design that combines RF signals across multiple satellites received using a low cost tiny receiver. We also describe the challenges in designing such a system and outline the techniques used to overcome those.

3.1 Design – Why RF Combination?

SelfieStick's approach to lower the cost and delay in accessing satellite images uses a tiny receiver module that receives signals from many Earth observation satellites. Although each received signal may have a slightly different underlying image, we preprocess the signals to transform and extract the common regions across the images. Post the transformations, we coherently combine the RF signals to recover the underlying image from noise that would otherwise have not been recoverable from a single weak reception.

Why not use traditional or deep learning based denoising? Denoising in image processing and computer vision filters noise in the image domain, either by assuming a particular distribution of noise or learning the distribution using a lot of data. However, our approach is driven by the fact the noise gets added in the RF domain due to weak signals received by our low-cost tiny receiver. The best way to filter the noise is by processing it in the RF domain. Instead, if one decodes the signal to image, the noise will get translated to the image domain through a non-linear digitization step. Further, it is well known that estimating phase measurements of highly noisy signals results in significant information loss [26] that makes signal recovery extremely challenging. This is mainly due to the "modulo 2π" aspect of signal phase that amplifies noise instead of useful signal at negative signal-to-noise ratio [26]. Learning a filter to de-noise such phase will require extensive training data for which there is no public RF dataset repository available, whereas RF processing requires only a few receptions.

Why not design cheaper or distributed ground stations? The cost of installation and operation of ground stations [42, 49] make it impractical to scale existing ground station infrastructure to cover the entire globe. An oft-ignored fact is that one needs to install the bulky ground stations at appropriate locations, away from structures that might occlude the signal from satellites. This is why ground stations are generally deployed on hills or rooftops. Therefore, the amount one would have to spend on installation might turn out to be more than that of the ground station's physical hardware. In contrast, we envision SelfieStick ground stations to be portable hand-held platforms. While using distributed receivers might seem a viable option, we seek to reduce the burden of cost and infrastructure on an individual further by transferring the diversity from the receiver to the transmitter. Distributed receivers require collaboration and technical know-how to deploy and generate an image, whereas our design could work standalone and cost a fraction of the multi-receiver approach. This could pave the way for satellite enthusiasts (researchers, amateur radio operators, etc.) as well as individuals in remote areas to access data from satellites that have already been launched by other large government organizations and private companies without incurring huge costs.

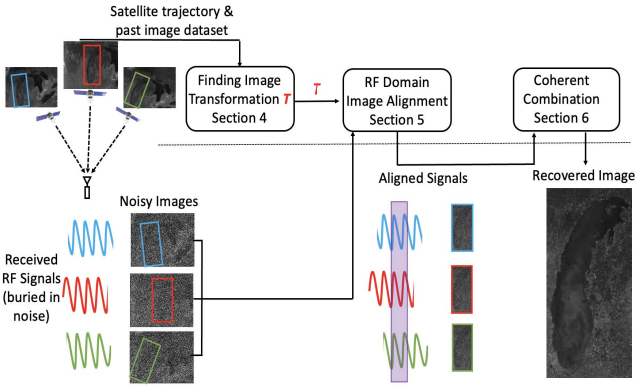

Figure 3: *SelfieStick's Architecture*: SelfieStick transforms received satellite images *indirectly* in the RF domain, prior to coherently combining these signals.

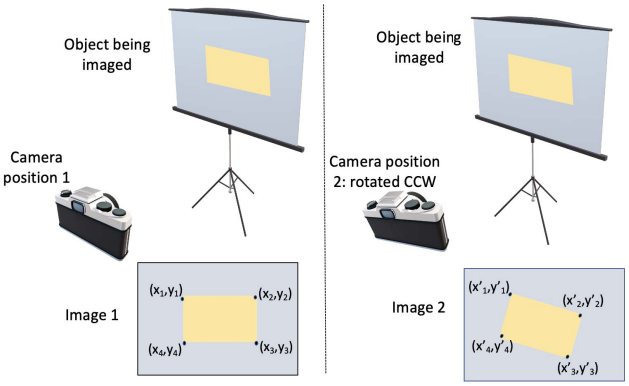

Figure 4: Difference in image captured due to rotation of camera in counterclockwise direction from left to right

3.2 Challenges and System Outline

The rest of this paper discusses three key challenges in SelfieStick's design. We illustrate how each of these challenges interplay to contribute to SelfieStick's architecture in Fig. 3.

Image Transformation: Since the satellites take pictures of the same scene from different locations and orientations, one cannot directly combine the received RF signals because the underlying images would appear to be taken from different perspectives. We must bring all the images to a common perspective if we are to successfully combine the RF signals. Sec. 4 discusses our approach to correct for these multiple perspective effects by learning a transformation matrix using a data driven approach based on the satellite's trajectory and camera sensor parameters.

Image Alignment in the RF Domain: Once the transformation is learnt, we cannot directly apply it to the image, since that would require decoding the RF samples to image in the first place. We instead use the reverse mapping from image pixels to RF samples to indirectly apply the transformation in the RF domain to mimic image transformation. Sec. 5 describes these two steps in detail.

Coherent Combination: Once the RF signals are appropriately transformed and aligned, we need to coherently combine them. To do so, we must first synchronize all the RF signals to get rid of any hardware dependent time, frequency and phase offset as well as trajectory dependent Doppler shifts. Sec. 6 describes our approach to overcome this challenge using the Sync markers present in the signal along with Doppler estimation using orbital parameters.

4 SATELLITE PERSPECTIVE ESTIMATION

In this section, we describe the problem of perspective distortion for LEO satellite images taken from different camera locations. We describe our solution to overcome this distortion in order to combine images correctly in the RF domain.

4.1 Perspective and Homography

Multi-view imaging is the concept of stitching together images taken from different perspectives to create a common combined view – an approach vital to the *panorama* mode of most smartphones [53]. The problem of multi-view is also very relevant to satellite imaging, especially for the case of LEO satellites which keep moving in their orbit with respect to the scene they image. A familiar example of this is Google Earth [53], where their consolidated continuous Earth images have been compiled by stitching together many small images after correcting for different perspectives.

Perspective in SelfieStick: In the context of SelfieStick, where the goal is to combine images from different satellites imaging the same broad region, we face the very same problem of different perspectives with a nuanced difference. In our case, the multiple perspectives are taken from different satellites and can therefore be fairly different, sometimes appearing rotated by almost 35 degrees with respect to one another. This perspective difference is actually more complex that just a rotation, often captured by effects like translation, rotation, skew, etc. In order to be able to combine images consistently from satellites at different locations and orientations, one needs to ensure that all the images appear to be taken from the same location and orientation. This is the problem of homography (see Fig. 4), which essentially tries to find a transformation (rotation, translation, shear, skew etc.) that converts one image to look similar to another image, given both the images are taken from vastly different perspectives. This is done through a reordering of the pixels in the image coordinate plane.

A Brief Primer on Homography: More formally, given a pixel at coordinates (x, y) in an image (see Fig. 4) and the corresponding same pixel at coordinates (x', y') in a second image, the homography matrix T is a 3x3 transformation matrix which when multiplied by coordinates (x', y') results in coordinates $(x, y) = T.(x', y')$. Note that both coordinate vectors (x, y) , (x', y') are appended by one to conform with matrix multiplication dimensions. The homography matrix T comprises of the combined effects of translation, rotation, scale, shear and perspective projections, which results in transformation of one coordinate to another. Given two satellite images, one can transform the second to look like the first image by estimating the transformation matrix. A straightforward way to estimate the transformation matrix is to find four corresponding points (x_i, y_i) and $(x_i', y_i'), i = 1, 2, 3, 4$ (no three of which are collinear) in the two images and solve for the eight unknowns elements of T using the eight equations (one of the elements is set to one, which only has a scaling effect).

4.2 Homography in SelfieStick Despite Noise

While the above formulation of homography works well if the corresponding points are known with high accuracy, this assumption fails in SelfieStick. Recall that signals received in our tiny RF receivers are weak, which will result in noisy images if decoded directly. Without a clean image, it is not possible to accurately find the four corresponding points needed for homography. Hence, we need to either find a way to accurately find the corresponding points without decoding the RF signals to images or find a different way to estimate the transformation matrix. Given this challenge, we go back to first principles to estimate the transformation matrix given the location of the camera (which depends on the trajectory of the satellite). To understand this, let us briefly recall how a camera generates a 2-D image of a 3-D scene [34] in the first place.

Satellite Imaging Viewed as Projections: The process of image generation is essentially a projection operation where the 3-D real world scene is projected onto a 2-D camera plane. Given a point (X, Y, Z) in 3-D space (see Fig. 5), its 2-D coordinates $\mathbf{x} = (x, y, 1)$ in the camera image plane can be found by multiplying the augmented 3-D point $\mathbf{X} = (X, Y, Z, 1)$ with a 3x4 camera matrix K such that $\mathbf{x} = K.\mathbf{X}$. Here, the points $\mathbf{X}$ and $\mathbf{x}$ are in the camera reference frame, with the camera placed at the origin. The camera matrix K captures camera intrinsic properties like focal length, skew, principal points and resolution scale factors in the x and y direction.

In our case, we operate in the Earth-Centered, Earth-Fixed (ECEF) reference frame where the 3-D coordinates of any point-of-interest relative to the Earth's center, assuming the Earth is fixed (i.e., no rotation) can be denoted as $\mathbf{X}$. Thus, the above camera projection equation gets modified to $\mathbf{x} = KR[I| - C]\mathbf{X}$. Here K is still the old intrinsic camera matrix, C is the camera location and R is a 3x3 matrix representing the orientation of the camera in the ECEF reference frame, I is a 3x3 identity matrix and $[I| - C]$ stands for column vector -$\mathbf{C}$ concatenated to the identity matrix I. From the above formulation, it is clear that if one knows the Camera matrix K, camera location C and Rotation matrix R, we can find the coordinates (x, y) in the camera image plane for any input 3-D point $\mathbf{X}$. In other words, should all these matrices be known, we can directly perform homography from first-principles.

The Problem of Satellite Orientation: The challenge however is, although K is fixed for a camera, and C can be found directly using the known trajectory of the satellite (available from satellite radar data or reverse-engineered from clean images previously collected), the matrix R can change. Recall that the R matrix characterizes the orientation (tumble) of the camera frame with respect to the ECEF reference frame as the satellite moves in space, and can be defined by three degrees of freedom: roll, pitch and yaw angles. Large Satellites have gyroscope and reaction wheels/thrusters for attitude control to appropriately align them to face the camera sensors towards the surface of the earth, which means the orientation angles in the ECEF frame vary based on the satellite's location [25]. These angles also depend on the trajectory of the satellite, initial deployment/calibration stage and factors like gravitational forces.

4.3 Data Driven Orientation Resolution

Our approach to estimate the orientation (tumble) of the satellite relies on prior clean images and the known satellite orbit to infer

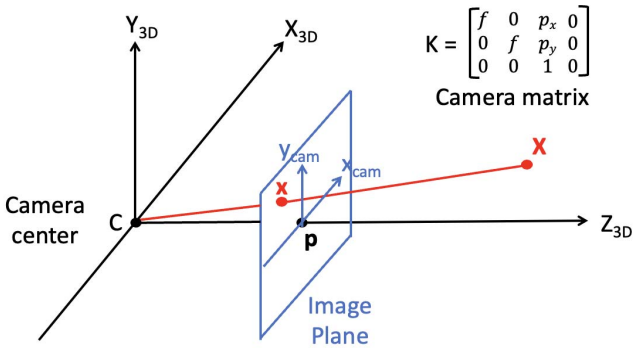

Figure 5: Imaging is essentially a projection defined by the camera matrix, which depends on the focal length and pixel position in the simplest form.

the evolution of its future orientation using a data-driven approach. This decision is driven by the fact that while onboard gyroscope data along with the satellite location would be sufficient to determine the orientation, it is hardly the case that data from gyroscope are sent downstream in decode-able format. It is directly used onboard for attitude correction. Even if this data was to be downlinked along the Tracking, Telemetry and Control (TTC) channel of the satellites in the future, there are two barriers to its adoption for SelfieStick: (1) The TTC channel is encrypted to ensure access only to the specific ground mission control, (2) TTC channel is transmitted on a different frequency band from the data downlink channel in all satellite missions. Thus, the receiver would need to be capable of tuning to multiple channels simultaneously using an addition RF chain (antenna, SDR), which would result in increased costs.

Why is orientation evolution predictable? To understand why this data driven approach works, recall that while the initial value of a satellite tumble is unpredictable when it is launched, it can be inferred once a clean image from the satellite is available at any point in its orbit, say when it is passing a major city. Should the orientation be known at a few initial points, its evolution in the future can be predicted with reasonable accuracy over long time scales (days). This is because air resistance and drag are negligible at LEO orbits and rotational velocity along any 3-D axis remains fairly consistent. Over longer time scales however, these assumptions may break down. Fortunately, most public satellites have large repositories of historical images and satellite orbital parameters [23, 31] allowing for a data-driven approach to make a regression-based estimate of future orientation.

Choice of Data-Driven Model: Thus, SelfieStick assumes that given the time series trajectory of the satellite and its corresponding orientation values at historical instances, we can learn a regression function for any future time instance. We train a multi-output deep learning regression model that uses five dense layers with ReLu activation units (see Fig. 6). The output layer returns the three parameters for roll, pitch and yaw although the last two parameters are mostly constant. The input features are the time series of the satellite locations at image capture and the corresponding camera intrinsic matrix. The training data is extracted from [31] which has a database of NOAA satellite images of different regions of the earth along with satellite locations at image capture. The output values

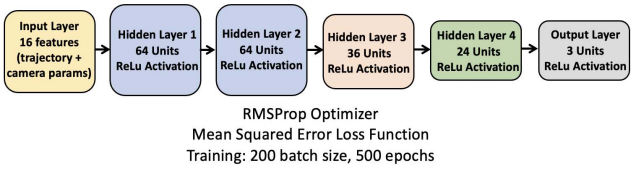

Figure 6: DNN Architecture

for the training data are generated by finding the rotation matrix (and hence the orientation parameters) from the clean images with respect to a reference image from the dataset. This is done by solving for the rotation matrix given the camera matrix for the reference image and the homography matrix between the reference image and the new candidate image as follows:

$$\begin{aligned} \mathbf{x} &= KR_{ref}[I|-C_{ref}]\mathbf{X} \\ \mathbf{x}' &= KR_{new}[I|-C_{new}]\mathbf{X} \\ \mathbf{x} &= T\mathbf{x}' \\ KR_{ref}[I|-C_{ref}] &= TKR_{new}[I|-C_{new}] \end{aligned} \tag{1}$$

Note that, here we can find the homography matrix T since both the reference image and new candidate image are clean. The loss function used for training the model is mean absolute error. Once the rotation parameters are estimated from the regression function, we can calculate the homography T matrix from the last equation in Eqn. 1 when the reference camera matrix and the new camera matrices along with the rotation matrices and satellite locations are known. It should be noted this method transforms all the images to the reference image's perspective.

Leveraging geographic information: With a data driven estimate of the homography matrix T in hand, one can further improve the estimate of the homography matrix by leveraging the local knowledge of prominent landforms in the region. Certain landforms can be used as distinguishing markers for a particular region, for example, the great lakes in North America. We create an RF signature of these landforms based on the specific modulation scheme used by the satellite. This is possible since past images of these prominent landmarks are readily available, and one just needs to modulate image pixels to RF samples. With an initial estimate of the homography matrix from the previous paragraph, we do gradient based search for the homography parameters that maximize the correlation with the RF signature of the landmark. Whenever these prominent landforms or their information is not available, we just use the data driven estimate learnt in the previous paragraph. In the next section, we use the homography matrix T to appropriately align the RF samples to mimic image transformation.

5 IMAGE ALIGNMENT IN THE RF DOMAIN

In this section, we seek to *align* received image pixels so that equivalent data across satellites are accurately combined. Specifically, given the transformation matrix T estimated in Sec. 4, the next logical step is to multiply each of the image matrix pixel indices with T to find the corresponding pixel indices in the transformed perspective-corrected reference frame. Once all the images are perspective corrected, one can add them coherently to boost the weak signal buried under noise. However, this would mean decoding the weak RF signals into image pixels in the first place. Should we attempt to decode weak RF samples to image pixels and then combine the image pixels, the resulting image would appear to be just noise, as explained previously in Sec. 3.1. Instead, SelfieStick seeks to indirectly apply the image transformation to the RF signals themselves, in a process we describe below.

5.1 Transformation in RF domain with BPSK

To find the appropriate pre-processing step in the RF domain that mimics the application of the transformation in the image domain, we need to formulate the following: (Q-1) What is the precise transformation applied in the image domain?; and (Q-2) What is the best mechanism to replicate this in the RF domain?

Defining the Transform in the Image Domain: To answer (Q-1) above, we draw from Sec. 4, where the application of the transform is essentially multiplication of the image pixel coordinates with the transformation matrix to generate the new pixel coordinate in the transformed reference frame. If $P^1_{x,y} = p$ is the value of the pixel at coordinates (x, y) in the reference frame #1 and (x', y') = $T.(x, y)$ are the transformed coordinates in reference frame #2 after multiplication with the transformation matrix[1], then the pixel value at coordinates (x', y') in the transformed reference frame #2 would be the same as that in frame #1, i.e, $P^2_{x',y'} = p$. Hence, we are translating the pixel from coordinates (x, y) in reference frame #1 to coordinates (x', y') in reference frame #2.

Replicating the Transform in the RF Domain: The answer to (Q-2) is more complex and depends on the modulation scheme employed to encode pixel values to RF samples during transmission. To understand this better, let us take a simple example (see Fig. 7) of a transmitter which employs the BPSK modulation scheme without coding. The camera sensor takes a photo of the scene and stores the 2-bit pixel values in an $M \times N$ image matrix. While transmitting the signal, it encodes the bit value of the pixels to BPSK symbols, shapes it onto a baseband pulse and modulates to a carrier frequency. The important point to note here is that the pixels are encoded into BPSK symbols in a line-by-line fashion, i.e., pixels in a row are first transmitted from left to right and then the pixels in the next row are transmitted. As a result, there is a direct correspondence between the pixel coordinates in the image domain and the RF sample number in the RF domain. In other words, any pixel at (x, y) in image #1 has a direct corresponding pixel (x', y') in image #2 at a deterministic RF time-series sample that can be directly computed. Fig. 7 shows an example of one such transformation. Note that it is quite often the case that (x', y') and (x, y) reside in completely different rows due to rotations or shear.

Since pairs of pixels in this representation have direct and deterministic correspondences, we can trivially find a deterministic permutation of RF samples that transforms the signal containing image #2 to image #1. The direct consequence of this is that in the transformed reference frame, the RF samples when properly aligned to mimic application of transformation, may comprise of non-contiguous time samples corresponding to different rows, since modulated pixel samples are transmitted in line-by-line fashion. We call this process RF sample *alignment*, where the goal is to

[1]In practice, (x', y') may not be integer coordinates. We deal with this in Sec. 5.3

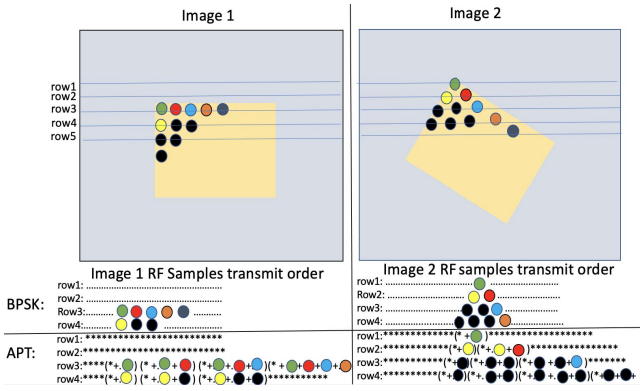

Figure 7: Pixels with the same color correspond to the same points in the 2 images. The corresponding RF samples are shown below for both BPSK and APT modulation. We can see that while the modulated pixel value is transmitted in BPSK, in APT modulation, each RF sample is equal to the cumulative sum of all previous samples

appropriately align the received time series RF samples along the transformed pixel coordinate time axis in non-contiguous time chunks to mimic image transformation.

Should the image pixel to RF modulation be as simple as the BPSK scheme above, our task is complete. However, satellite image modulations even from public satellites are much more complex. As a case study, we consider the popular NOAA satellites whose signals are modulated in APT format [30] in the next section. We discuss extending our approach to other modulations in Sec. 9.

5.2 SelfieStick with APT Modulation

In this section, we study NOAA satellites, mainly given that their modulation and decoding processes are openly available compared to other satellites where this information is proprietary or not publicly disclosed. We first briefly describe this modulation and how SelfieStick can be extended to apply to it.

FM Modulation in APT signals: Many of the small satellites in the LEO orbit employ Frequency based modulation schemes like FM, FSK and AFSK [47]. This is mainly due to their narrowband nature and robustness to noise in high attenuation scenarios usually encountered in satellite communication. As described in [30], NOAA APT signals employ a similar strategy of encoding their pixel values using AM modulation on a sinusoidal baseband carrier f_m which is then FM modulated. Recall that in FM modulation, the instantaneous frequency of the signal is modulated using the input baseband signal. The signal model is described in Eqn. 2 below

$$\begin{aligned} x_b(t) &= A(t)\cos(2\pi f_m t) \\ x(t) &= A_c e^{(j2\pi f_c t + j2\pi f_\Delta \int_0^t x_b(\tau)d\tau)} \end{aligned} \tag{2}$$

where the pixel value $A(t)$ is modulated on a sinusoid carrier f_m to generate a baseband signal $x_b(t)$. This baseband signal is then FM modulated onto an FM carrier f_c with frequency deviation f_Δ, which represents the maximum shift away from the carrier frequency f_c experienced by the FM modulated signal. It can be seen from the integral in the above formulation that the RF signal at any time is dependent on all of the previous samples. In our case, this can be rephrased as, the RF sample for a pixel value at any time is dependent on *all* the preceding pixel values in that line.

The Challenge with APT Signals: The dependence of each sample on all previous pixel values makes the RF sample alignment problem difficult compared to the BPSK example described above. It is especially challenging when the transformation requires aligning RF samples corresponding to pixel lying in different rows. This is because in the BPSK case, one could just pick RF samples corresponding to a particular pixel across different images and add them in a phase coherent manner to recover the signal from noise. The inherent assumption that makes this work is the fact that the underlying RF signal for any particular pixel value across images is the same, modulo noise. However, in case of FM modulation, this assumption breaks down since the underlying RF signal corresponding to the same pixel value across different images is no longer the same, it is dependent on the previous pixel values. Crucially, these can be different, especially in the case when the images are transformed versions of each other (see Fig. 7 for an example). Thus, we need a way to remove this dependence on the previous pixel values if we were to carry out the alignment process correctly.

RF Alignment for APT signals: To remove the dependence of previous pixel values on the current pixel RF samples, we leverage the fact that this dependence can be characterized as a linear function of previous pixel values. Specifically, from Eqn. 2 it follows that the instantaneous frequency of an APT modulated signal after removing the carrier f_c post passband demodulation, is the integral of the AM modulated pixel values from previous time stamps. This can be understood as the current pixel RF sample value being equal to the exponential of the sum of all the previous pixel values up to (and including) the current pixel value. In other words, the phase of an RF sample is directly proportional to the phase of the previous RF sample plus the contribution from the current pixel. Hence, to extract a single desired pixel value, one can simply subtract the phases of adjacent RF samples. In the complex domain of RF samples, subtraction can be performed using multiplication by the conjugate, which we describe below:

$$\begin{aligned} x_{ind}(t) &= x(t) * x^*(t_i)/|x(t_{i-1})|^2 \\ &= e^{(j2\pi f_\Delta \int_{t_{i-1}}^{t_i} x_b(\tau)d\tau)} \end{aligned} \tag{3}$$

where t_i is the time index corresponding the i^{th} pixel's RF sample. This correction is done for the RF samples corresponding to all the pixels across the different images at the time of reception to generate a RF sequence independent of past samples. Note that this step is necessary to ensure correct RF alignment. Upon this correction, each RF sample is solely dependent on one pixel. After this step, one could carry out the alignment exactly as described in the BPSK example above from Sec. 5.1. Once all the samples of i^{th} receiver is aligned to a common reference frame, it is ready for coherent combination. Notationally, we say that RF samples from satellite i: $x_{i,ind}(t)$ have been aligned to become $x_{i,ind}^{aligned}(t)$

5.3 Handling Alignment Non-Idealities

Next, we describe a few sources of non-idealities in our alignment approach thus far: fractional pixel locations and signal noise.

Fractional Pixel Coordinates: Naïve application of transformation described earlier might result in pixel coordinates that are fractional in the transformed reference frame. This is because, application of transform is essentially multiplication of the transformation matrix with pixel coordinates: $(x', y') = T.(x, y)$. This would mean that either x' or y' or both could compute to a fractional value. One straightforward way to handle this is to round the fractional values to the closest whole number and assign the pixel value p to the rounded pixel coordinates $round(x', y')$. However, this can result in multiple pixel values as candidates for a particular coordinate since there might be two coordinates (x_1, y_1) and (x_2, y_2) that will result in the same transformed pixel coordinate (x', y').

To overcome this problem of fractional pixel coordinates, we use the inverse transformation matrix to go from the transformed reference frame to the original reference frame: $(x, y) = T^{-1}.(x', y')$. This ensures that the pixels in the transformed reference frame are always whole numbers, but the pixels in the original reference frame can now be fractional. The advantage of this approach is that since we have the RF samples corresponding to the pixel values in the original frame at any integer coordinates, we can interpolate these values to estimate what would have been the RF sample corresponding to fractional coordinates. We use bilinear interpolation to estimate the pixel value at any fractional coordinate in the original reference frame. Bilinear interpolation is a classical technique that takes an appropriate weighted average of neighboring integer coordinate pixels to estimate pixel value at a fractional coordinate. Given that pixel values are encoded in the phase of RF samples, we apply this interpolation – a weighted sum of pixels – in the RF domain as an exponentiated product of samples.

Mathematically, suppose (x, y) computed from the inverse transformation matrix is a fractional value. Then we find the bilinear weights k_1,k_2,k_3 and k_4 corresponding to the pixels p_1,p_2,p_3 and p_4 at coordinates $(\lfloor x \rfloor, \lfloor y \rfloor)$, $(\lfloor x \rfloor, \lceil y \rceil)$, $(\lceil x \rceil, \lfloor y \rfloor)$ and $(\lceil x \rceil, \lceil y \rceil)$ respectively. To find the RF samples corresponding to (x', y'), we apply the bilinear transformation in the RF domain as:

$$\begin{aligned} x_{ind}(t_{x',y'}) = {} & x_{ind}(t_{\lfloor x \rfloor, \lfloor y \rfloor})^{k_1} x_{ind}(t_{\lfloor x \rfloor, \lceil y \rceil})^{k_2} \\ & \times x_{ind}(t_{\lceil x \rceil, \lfloor y \rfloor})^{k_3} x_{ind}(t_{\lceil x \rceil, \lceil y \rceil})^{k_4} \end{aligned} \tag{4}$$

$$\text{where: } x_{ind}(t_{x,y}) = e^{(j2\pi f_\Delta \int_{t_{x,y-1}}^{t_{x,y}} p_1 cos(2\pi f_m \tau) d\tau)} \tag{5}$$

which is the RF sample corresponding to the pixel at coordinates (x, y). The process in Eqn. 4 is equivalent to finding the bilinearly interpolated pixel value at fractional coordinates (x, y) and assigning it to the coordinate (x', y') in the transformed domain.

Effect of Noise: The operations in Eqn. 3 and 4 can be designed to be robust to noise. For the operation in Eqn. 3 to work in the presence of noise, we average the oversampled RF samples across the vicinity of the current pixels. This helps to reduce the effect of noise, since adjacent pixels in a satellite image are almost similar in values, unless it is a boundary pixel. The application of exponent k_i in equation 4 also does not accentuate noise, since the values of the k_i weights are always less than 1.

6 COHERENT MULTI-VIEW COMBINATION

In this section, we describe the key synchronization steps that need to be performed, as well as the steps to actively combine signal receptions across satellites.

6.1 Channel Model and Non-Idealities

Prior to combining RF signals from across many diverse satellites, it is necessary that all the signals are synchronized in time, frequency and phase. Synchronization is essential in the context of our multi-transmitter coherent combination system because: (1) It is not possible for the satellites transmitters and our tiny receiver to share the a common clock (GPS clock sharing is one of the popular solutions, but it increases the cost and doesn't work well in indoor deployments), (2) Signals from LEO satellites are much adversely affected by a time varying frequency offset called Doppler shift due to their high orbital speeds and (3) there is no way for the users to communicate with the satellite (except for users belonging to the organization operating the satellites), and hence, synchronization of our tiny receiver to the satellite through two-way communication is not an option. Since we are receiving signals from many different satellites, it is essential to remove offsets between each satellite and the receiver before combining the multiple signals.

Characterizing Hardware and Doppler offsets: We consider a narrowband signal model to characterize the signal received by our tiny receiver from satellite i. The signal is affected by 3 types of offsets: (1) Carrier Frequency Offsets: when the receiver's carrier frequency (f) is different from that of the satellite (f_i), (2) Sampling Frequency Offsets: when the receiver's sampling time (t) is different from that of the satellite (t_i) and (3) Phase Offsets: when the locking of the receiver's phase locked loop locks at a phase (ϕ) that is different from the satellite (ϕ_i). Doppler shift (f_{d_i}) is another frequency offset specific to LEO satellites caused due their high orbital speeds, and must be corrected for, since it can reach values of over 1 KHz across a one minute reception period. More formally, our signal model is:

$$y_i(t) = h_i e^{j(2\pi f(t-t_i)+2\pi(f-f_{d_i}-f_i)t+(\phi-\phi_i))} x_i(t) \tag{6}$$

Where y_i is the received signal from satellite i, h_i is the true offset free channel between the satellite and the receiver, the exponential term is the cumulative effect of the hardware and trajectory induced offsets. Next, we describe how we nullify these offsets in order to synchronize the receivers to each of the satellites.

6.2 Correction and Coherent Combination

Doppler Correction: We first correct for Doppler offsets experienced by our receiver from different satellites in orbit. The Doppler shift is based on the geographical position of the receiver and the satellite's trajectory, which make it different for different satellites, since the satellite's trajectory is determined by its Keplerian orbital parameters. The Simplified Perturbations Model (SGP4) [33] propagates the satellite's trajectory based on the input orbital parameters through Two Line Elements (TLE) and generates its location and velocity at any given point in time. Another parameter that can be computed from the SGP4 model is the Doppler shift with respect to any point on Earth based on the transceivers center frequency and the range rate. We can remove the Doppler offset from our received signal due to the i^{th} satellite by multiplying $e^{j2\pi\hat{f}_{d_i}t}$ with the measured channel, where $\hat{f}_{d_i}$ is the Doppler shift predicted by the SGP4 model.

Hardware Offset Correction: Although the signal received is buried under noise, we can leverage NOAA's Synchronization Markers A and B [30] to estimate and nullify the effect of the remaining hardware offsets in the signal. This is because the sync markers, like preambles used in many communication systems are long and structured sequences that can be detected at lower SNRs with correlation even if the data payload is buried under noise. Once the preamble is detected in the received signal, its phase variation over time and frequency can be used to estimate the first two hardware offsets. In particular, the carrier frequency offset term $(f - f_i)$ appears a linear variation of phase over time. Hence, by measuring the slope of the averaged phase over time of the sync marker in the received signal, the carrier frequency offset can be estimated. Similarly, the sampling offset term $(t - t_i)$ manifests as a linear variation of phase over frequency when the signal is viewed in the frequency domain. We measure the slope in the frequency domain to estimate the sampling frequency offset. The phase offset $(\phi - \phi_i)$ term remains constant over time as well as frequency, and hence, can be estimated using a simple one tap equalizer.

Coherent combination: The next step after synchronization is the channel equalization step, where we remove the effect of channel h_i by multiplying with its conjugate h_i^*. This has two benefits: (1) We do not need to track the evolution of the channel term as the RF sample goes through the different steps of alignment, and (2) We already have the contribution of the sample $x_{i,ind}(t)$ for coherent combination where $x_{i,ind}^{aligned}(t)$ is the corresponding aligned RF sample from satellite i (see Sec. 5). Thus, if N such aligned samples are available from N satellites, we can combine them as:

$$x_{\text{combined}} = \sum_{i=1}^{N} x_{i,ind}^{aligned}(t_{x',y'}) \tag{7}$$

With this process, each satellite's aligned RF sample adds up coherently while the noise being random, adds up incoherently. This provides a gain in SNR compared to a single satellite case, which is known as the diversity gain.

Putting it all together, SelfieStick proceeds in three steps: (1) First synchronization to eliminate hardware offsets and doppler shifts are performed (Sec. 6.1 above); (2) Second, image alignment is performed as described in Sec. 5 based on the transforms from Sec. 4. (3) Finally, the coherent combination step described above.

7 IMPLEMENTATION

Hardware: SelfieStick's receiver setup comprises of an antenna, a software defined radio (SDR) and a computer. At the antenna front, we use the Laird Technologies EXS136SMI Tuf Duck antenna, which offers unity gain with toroidal radiation pattern and vertical polarization. It is rated to operate between 136 MHz and 150 MHz while connected to handheld VHF radio receivers like ICOM F50, etc. Its operating frequency overlaps with the US NOAA satellites APT transmission at 137 MHz, which made it a perfect candidate for our low-cost tiny receiver. On the SDR front, we use the popular low-cost RTL-SDR dongle R820T2 RTL2832U. Its low cost along with the ease of use made it a natural choice for our design. The total cost of our RF-frontend (including the SDR) is $38 with a form-factor as small as 20 by 3 centimeters. A Raspberry Pi system with a USB port or a laptop can be used to plug in the RTL-SDR.

Software and Testbed: The reception parameters like sampling rate and center frequency are set up using RTL-SDR software installed on the computer. To predict when to collect the data, we use the SGP4 propagator with the orbital elements TLE file for the satellites to predict the time of satellite's pass overhead. These TLE files are made publicly available by NORAD on a periodic basis. Once sufficient satellite data is collected, it can be preprocessed locally to correct for perspective differences before coherent combination to generate the underlying image. We use Python and MATLAB for all the Machine Learning and Digital Signal processing tasks described in Sec. 4, 5 and 6 as well as for generating the satellite image. We evaluate our system both indoors (inside buildings) and outdoors (on rooftops of buildings) in a university campus. To compare the performance of our system against clean images, we install the UC-1374-531R quadrifilar helix weather antenna [3] (see Fig. 8A) on a campus building rooftop, connected to the LNAU-0137-648 low noise amplifier and the USRP N210 SDR to collect the clean reference images used in Sec. 8.3. This bulky antenna provides a gain of 10-12 dB compared to our unity gain low form-factor Tuf-Duck antenna, which results in received SNR difference of 10-15 dB, depending on the location of deployment.

8 RESULTS

We evaluate the performance of SelfieStick for different microbenchmarks and performance metrics. As mentioned earlier, we design our system with the NOAA case study, and hence in all our experiments we receive 4 km per pixel low resolution image signals from NOAA 15, NOAA 18 and NOAA 19 satellites operating in the 137 MHz VHF band. These satellites orbit the Earth at an altitude of 850 km, with one of their transmitters continuously emitting APT [30] modulated signals with real-time images of Earth under them. This choice of NOAA satellites was driven by the availability of hardware as well as public information about the transmission scheme and packet structure. Since, today there are not enough NOAA satellites that appear over a particular region of Earth simultaneously, we emulate multi-satellite case, by combining signals received from these at different points in time. The key insight that makes this model work is that regions of Earth (features) from space appear the same over long periods of day modulo cloud cover and perspective distortion. Hence, with each NOAA LEO satellite passing almost 3-5 times over any non-polar region of the Earth, we have 5-6 opportunities to receive signals each during the day and night passes. We also combine signals across 2-3 days in cases with clear sky to emulate the larger transmitter diversity scenarios.

8.1 Microbenchmarks

Satellite Camera Orientation: As described in Sec. 4, we need to bring all received images to a common perspective in order to combine them effectively. To estimate the transformation needed on each of the images indirectly, we calculate the satellite internal camera parameters and combine it with the estimate of satellites location (SGP4 model) and camera orientation (deep learning model). Here, we characterize the accuracy of satellite camera roll angle prediction of our deep learning models trained for the three different satellites mentioned above (model details and dataset information

Figure 8: Hardware used in SelfieStick (A) Weather Antenna, (B) RTL SDR with Laird Antenna

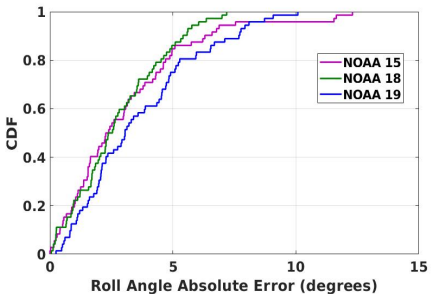

Figure 9: CDF of Roll Angle Absolute error

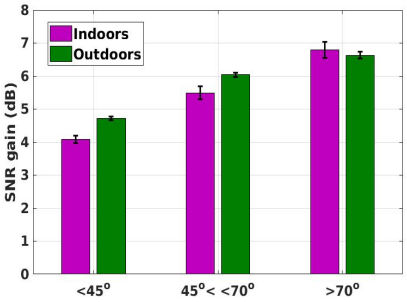

Figure 10: SNR gain at different maximum elevation angles

detailed in Sec.4). Fig. 9 shows the CDF of the absolute error in prediction of the camera roll angle from the true roll angle. It can be seen that our model achieves median absolute error of 2.3 degrees, 2.5 degrees and 3.1 degrees for NOAA 15, NOAA 18 and NOAA 19 respectively. The effect of this error is negligible because of two reasons: (1) the elements of the roll rotation matrix essentially comprise of the sine and the cosine of the angle, which changes slowly; and (2) the distance between the satellite camera and the scene imaged is huge, which make the image captured less affected by these minute differences in rotation angle.

Latency: As mentioned earlier, our evaluation is based on temporally separated data collected from three NOAA satellites which requires almost 11 hours to collect 10 distinct data samples. This is due to the coupled effect of Earth's rotation and the polar orbit of the NOAA LEO satellites which makes them appear over any non-polar region only 2-3 times during the day and then again after a gap of 11-12 hours. However, this latency can be greatly reduced by increasing the constellation size. To motivate this, we use the trajectory of all satellites from 2 Earth observation constellations currently in orbit: Planet Labs (189 satellites), and Spire constellations (120 satellites). We find that Planet Labs entails a median latency of 23 minutes compared to 36 minutes for the Spire constellation in receiving signals from 10 satellites. With the exponential trend [11] in the number of LEO satellites being launched today, this latency will inevitably drop to seconds in the near future.

8.2 RF metrics

Transmit Diversity gain: To characterize the average SNR gain in SelfieStick, we collect satellite signals from ten temporally separated satellite passes from receivers placed both indoors and outdoors. We then coherently combine them after perspective correction and synchronization. The SNR gain is calculated each time after adding additional satellite reception with respect to the average SNR over the 10 receptions as the baseline single satellite reception case. The average SNR gain increases logarithmically with the number of satellite receptions used for coherent combination for both indoor and outdoor settings. Fig. 11 shows that SelfieStick achieves an average SNR gain of 8.4 dB using 10 satellite receptions.

Satellite Elevation Angle: The elevation angle of a satellite with respect to a ground station on Earth is the angle at which one must point a directional antenna to receive signals from the satellites. Large propagation loss and obstruction from nearby ground infrastructure cause the received signal strength to drop significantly at low elevation angles. In the context of SelfieStick low elevation angles have the added disadvantage of increased amount of perspective distortion in the captures image. We analyze the effect that elevation angle has on SelfieStick with the following experiment: Instead of combining signals across all elevation angles, we only combine signals received from passes whose maximum elevation angles fall within the following three categories: (1) maximum elevation angle below 45 degrees, (2) maximum elevation angle from 45 to 70 degrees and (3) maximum elevation over 70 degrees. We then plot the average SNR gain achieved in the 3 categories for 6 satellite passes in Fig. 10. We can see that low elevation angle passes indeed suffer from lower gains. The gap in SNR gain achieved between indoor and outdoor settings is also larger for low elevation angles due to higher attenuation of indoor signals from building walls.

8.3 Image metrics

Since we are dealing with satellite images, one must check if the SNR gain achieved in the previous subsection suffices for image decoding. We use 2 popular metrics used to quantify goodness of SelfieStick's recovered image with respect to a reference image: Structural Similarity Index Measure (SSIM) [48] and Peak Signal to Noise Ratio (PSNR) [20]. The reference image used for this comparison is captured by the Weather antenna (see Fig. 8A) placed on the roof of a university campus building as described in Sec. 7. For both these metrics, we run the same experiment as that from Sec. 8.3, except now we also generate the image after adding every additional satellite reception. We then compute the SSIM and PSNR metrics of the generated image with the reference image.

SSIM Variation: The Structural Similarity index compares two images based on the following parameters: structure, contrast and luminance. The SSIM value lies between 0 and 1, with values close to 1 indicating a good match between the 2 images. We can see from Fig. 12 that as we move from one satellite reception to 10 satellite receptions, the average SSIM improves from 0.53 to 0.83 for outdoor settings. Average SSIM also improves for indoor settings, however its absolute value is less than that achieved outdoor, which is expected given higher signal attenuation experienced indoors.

PSNR Variation: The PSNR metric compares 2 images by measuring the mean squared error of corresponding pixel values across the

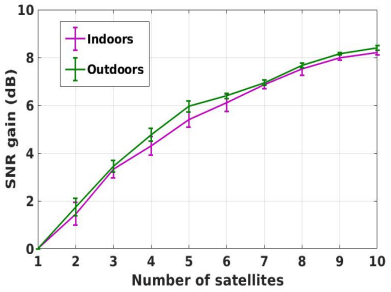

Figure 11: SNR gain with increasing number of satellites

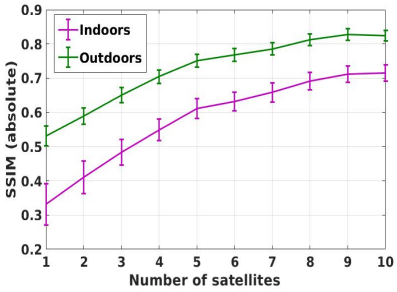

Figure 12: SSIM variation with increasing number of satellites

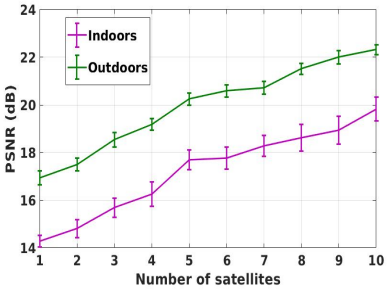

Figure 13: PSNR improvement with increasing number of satellites

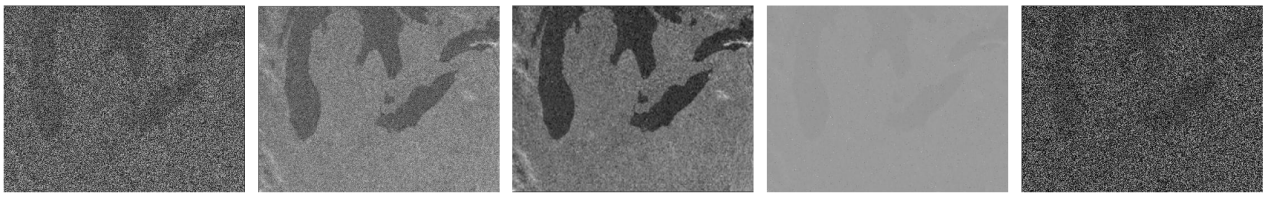

Figure 14: Image produced by (a) SelfieStick1 Satellite, (b) SelfieStick10 Satellites, (c) Bulky Weather Antenna, (d) DNN based Denoiser, (e) BM3D based Denoiser

image and returns its inverse. We can see that SelfieStick achieves an average 5 dB improvement in PSNR as one moves from a single satellite reception to 10 satellites reception from Fig. 13.

Comparison with Image Denoising: We compare our system with Block Matching and 3-D filtering (BM3D)[13] and MATLAB's DNN based denoiser. Fig. 14(c) plots the zoomed in image of the Great Lakes received from the bulky Weather antenna in Fig. 8A. We compare it with that generated by BM3D filtering (Fig. 14(e)), DNN Denoiser (Fig. 14(d)), SelfieStick with single reception (Fig. 14(a)) and SelfieStick with 10 receptions coherently combined (Fig. 14(b)). We see that although the recovered image in SelfieStick doesn't achieve the exact match with the weather antenna image, it still does much better than single reception, DNN and BM3D.

9 DISCUSSION AND LIMITATIONS

Generalizing to other Modulations: Our discussion in this paper so far is based on APT modulation, however the techniques are applicable to any modulation scheme that modulates pixel values directly into RF samples (either in its phase, frequency or amplitude). Another common modulation scheme used by weather satellites like NOAA, METEOR and METOP where our techniques are applicable are Low Resolution Picture Transmission (LRPT) [39] and High Resolution Picture Transmission (HRPT) [32] where QPSK based phase modulation is used to encode pixels values and transmitted at a much higher rate. Note that in LRPT and HRPT, the problem is quite similar to the BPSK case described in Sec. 5.1 due to an intuitive mapping between pixels and I/Q samples. For more complicated encoding schemes like JPEG 2000 used in satellites like Sentinel constellation, our techniques will not directly apply due to the wavelet transform based step followed by Huffmann coding. To overcome this hurdle, one could build a more complex learning model that maps pixels to RF samples, however that is beyond the scope of this paper. The information about the modulation and encoding scheme used by private satellites is proprietary which reduces our focus to only government owned satellites.

Currently, our techniques depend on the RF samples being modulated with the same scheme. However, in case a different modulation scheme is used, one would have to demodulate to the pixel level RF samples to combine in the RF domain. However, having the same modulation scheme across many satellites is a realistic assumption since many of the earth observation satellites being launched today are parts of large constellations like PlanetLabs, NOAA POES, SENTINEL. Ideally, one would want a consistent modulation format standardized like DVB for satellite TV that could be used for multiple satellites for earth observation. This is indeed true for some cases like the HRPT modulation format used across different space agencies like NASA (for NOAA satellites) and ESA (for METEOR satellites). One of the main advantages of using a standardized modulation format is that one does not have to deal with resampling, resizing, and initial demodulation at vastly separated frequencies before SelfieStick's techniques can be applied.

Effect of cloud cover: A key assumption behind our system is that the underlying scene remains the same across time. However, this might not be true, especially with heavy cloud cover. Since, our receptions are now distributed across time, the underlying cloud cover might change over that period of time, impeding SelfieStick. However, many satellites transmit multi-spectral images of the scene (including some of the newer NOAA satellites), which may capture images bereft of cloud occlusions that SelfieStick can leverage. Note that the techniques would still be applicable in cloudy scenarios once there are a sufficient number of satellites over a region transmitting simultaneously to ensure image recovery, however, the image decoded would still contain the clouds.

Applicability to Large Images: As long as the images captured are of the same scene (from different perspectives) as well as the modulation scheme is known, we can apply SelfieStick to recover the image. Large-scale high-resolution images like those produced by Wide Area Motion Imagery (WAMI) [19], are generated by stitching together multiple high-resolution images of smaller scenes to create a composite larger scene, either on board the aircraft or by downlinking the individual smaller images to a ground station. SelfieStick's techniques can be applied on the raw RF samples all at once on the composite image or on the individual smaller chunks depending on the onboard processing. However, due to the massive size of high resolution images, SelfieStick would need to deal with a much larger number of RF samples (corresponding to individual pixels), something that could perhaps be offloaded to cloud. One degrading factor could be the motion of individual objects like vehicles, if the images are taken far apart in time. This could cause blurring in the final image produced. However, WAMI being motion imagery is robust to this since the frames are captured extremely quickly to generate a motion image.

Bootstrapping, Scalability and Long-term Evolution: To be successful, SelfieStick does require periodic orbital parameters and historical image datasets from the satellite. Ground stations must have Internet connectivity to pull satellite orbital parameters. SelfieStick's PSNR gains are dependent on the number of overhead satellites at any time, which can vary with geography and depending on frequency of operation – we refer the reader to Sec. 8.3 for a PSNR vs. scale analysis. We believe SelfieStick's techniques and performance would scale as the LEO technology improves and matures in the future leading to more and more satellites being launched. As LEO satellites are miniaturized, their power budgets become smaller and smaller, which directly influences the transmit power. As a result, one would require more bulkier antennas to receive even weaker signals. Our techniques become even more valuable in this case since we leverage the diversity of the increasing number of satellites to receive weak signals using a low-cost module.

10 RELATED WORK

Multiview and Image Denoising: There exists a huge body of work on multiview geometry [18, 34, 45, 51, 53], with applications ranging from photography to Satellite imaging with the broader goal of enhancing the scene captured by stitching together multiple images after rectification. While we use the basic principles of multiview geometry and camera projections to build our system, we differentiate ourselves from these in terms of the way we apply the multiview transformations on the RF samples to mimic image transformation. Application of classical image perspective correction techniques like SIFT and SURF on our noisy images, result in large mismatch in the top 50 corresponding points returned, due to the presence of multiple similar intensity regions in grayscale images further degraded by significant noise. Similarly, satellite imagery is one of the major applications for image denoising in the computer vision community, employing a combination of wavelet as well as deep learning based techniques [10, 17, 21, 43] to estimate the underlying distribution of noise and attempt to filter it. What differentiates us is that we are trying to denoise the images by dealing with it in the RF domain, while requiring significantly less data. This is essential for our case since signals received by our tiny receivers are buried under noise which makes it difficult to filter it once decoded into images; and lack of access to large amounts of satellite RF dataset that could be used for ML based denoising. We do train a Deep learning model that learns the orientation parameters of the satellite, however this done using clean images from past receptions and is limited to this purpose only - not denoising. We compared our system with both wavelet and deep learning based filters in Sec. 8.3, showing the efficacy of our techniques.

Satellite Imagery and RF based techniques: There has been a lot of interest in recent years in accessing satellite data with low latency, mainly driven by the exponential increase in the number of satellites being launched. Recent efforts towards this goal involves either increasing the number and geographic distribution of ground stations [28, 46], either by using community based ground station networks [37, 44] or commercial entities renting out their ground stations[2, 6]. While these have potential to improve the access to satellite data, there is still a gap in acceptance of these approaches especially in remote regions and developing nations.

Using RF signals to enhance imaging applications has been widely used in both medical and space domains, with ultrasonic imaging and Synthetic Aperture Radar techniques [9, 12, 38, 41]. However, we differentiate ourselves from these in the sense that we use the very weak RF signals to recover the images, whereas most of the SAR based techniques require directional and high-power signals to generate the image.

Coherent combination: Coherent combination has been widely used in wireless systems to improve communication [14, 16, 40] and sensing applications[24, 29, 35]. Transmit diversity techniques have also been studied for MIMO and multi antennas systems, but we focus on its application in satellite context. There have been some proposals to use transmit diversity for satellite systems especially in broadcast applications [5, 27, 52]. We differ from these in attempting to combine satellite signals containing differing views of the Earth.

11 CONCLUSION AND FUTURE WORK

We present SelfieStick, a new satellite ground station system that can provide real-time access to satellite images using a tiny low-cost receiver module that can be deployed both indoors and outdoors. We do so by coherently combining weak signals received from multiple LEO satellites, each transmitting slightly different views of the underlying scene. We develop techniques to convert these views to a common view in the RF domain, a necessary step before coherent combination that boosts the signal to noise ratio. We believe this work will be useful in overcoming the cost and latency barrier associated with satellite data access faced by users in remote areas. As next steps, we hope to extend SelfieStick to a wide range of frequencies of operation. We will also explore rich applications atop low-cost satellite imagery.

Acknowledgements: We thank the shepherd and reviewers for their insightful feedback which helped improve this paper. We thank Vijay Rengarajan and Akarsh Prabhakara for invaluable discussions during different phases of the project. We also thank all the members of the WiTech Lab at CMU for their inputs. This research was supported in parts by NSF grants 2030154, 1942902 and 2007786, Kavčić-Moura grant and IoT@CyLab.

REFERENCES

[1] 2020. Commercial Satellites. Retrieved April 30, 2020 from http://gsp.humboldt.edu/OLM/Courses/GSP_216_Online/lesson3-2/commercial.html

[2] 2020. Infostellar. https://infostellar.net/

[3] 2020. WXSAT-APT-KIT Satellite Weather Kit. Retrieved October 20, 2020 from https://www.antennas.us/WXSAT-APT-KIT-Satellite-Weather-Kit.html

[4] 2021. Planet Labs. Retrieved October 29, 2021 from https://www.planet.com/

[5] Do Seob Ahn, Sooyoung Kim, Hee Wook Kim, and Dong-Chul Park. 2010. A cooperative transmit diversity scheme for mobile satellite broadcasting systems. *International Journal of Satellite Communications and Networking* 28, 5-6 (2010), 352–368.

[6] amazon.com. 2020. AWS ground stations. Retrieved August 21,2020 from https://aws.amazon.com/ground-station/

[7] Simon I. Hay Andrew J. Tatem, Scott J. Goetz. 2020. Fifty Years of Earth-observation Satellites. Retrieved April 30, 2020 from https://www.americanscientist.org/article/fifty-years-of-earth-observation-satellites

[8] Tor Aulin, Nils Rydbeck, and C-E Sundberg. 1981. Continuous phase modulation-Part II: Partial response signaling. *IEEE Transactions on Communications* 29, 3 (1981), 210–225.

[9] Paola Benedetti, Dino Ienco, Raffaele Gaetano, Kenji Ose, Ruggero G Pensa, and Stephane Dupuy. 2018. M3Fusion: A Deep Learning Architecture for Multiscale Multimodal Multitemporal Satellite Data Fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11, 12 (2018), 4939–4949.

[10] Ashish Kumar Bhandari, D Kumar, Anil Kumar, and Girish Kumar Singh. 2016. Optimal sub-band adaptive thresholding based edge preserved satellite image denoising using adaptive differential evolution algorithm. *Neurocomputing* 174 (2016), 698–721.

[11] Supriya Chakrabarti. 2021. How many satellites are orbiting Earth? Retrieved October 20, 2021 from https://astronomy.com/news/2021/09/how-many-satellites-are-orbiting-earth

[12] John C Curlander and Robert N McDonough. 1991. *Synthetic aperture radar*. Vol. 11. Wiley, New York.

[13] Aram Danielyan, Vladimir Katkovnik, and Karen Egiazarian. 2011. BM3D frames and variational image deblurring. *IEEE Transactions on image processing* 21, 4 (2011), 1715–1728.

[14] Adwait Dongare, Revathy Narayanan, Akshay Gadre, Anh Luong, Artur Balanuta, Swarun Kumar, Bob Iannucci, and Anthony Rowe. 2018. Charm: exploiting geographical diversity through coherent combining in low-power wide-area networks. In *2018 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 60–71.

[15] The Antenna Farm. 2021. Laird Technologies EXS-136-SMI. Retrieved October 21, 2021 from https://www.theantennafarm.com/catalog/laird-technologies-exs-136-smi-2288

[16] GD Goodno, H Komine, SJ McNaught, SB Weiss, S Redmond, W Long, R Simpson, EC Cheung, D Howland, P Epp, et al. 2006. Coherent combination of high-power, zigzag slab lasers. *Optics letters* 31, 9 (2006), 1247–1249.

[17] Bhawna Goyal, Ayush Dogra, Sunil Agrawal, Balwinder Singh Sohi, and Apoorav Sharma. 2020. Image denoising review: From classical to state-of-the-art approaches. *Information fusion* 55 (2020), 220–244.

[18] Rostam Affendi Hamzah and Haidi Ibrahim. 2016. Literature survey on stereo vision disparity map algorithms. *Journal of Sensors* 2016 (2016).

[19] BAE Systems Inc. 2022. What is Wide Area Motion Imagery? Retrieved February 10, 2022 from https://www.baesystems.com/en-us/definition/what-is-wide-area-motion-imagery#:~:text=Wide%20Area%20Motion%20Imagery%20(WAMI,within%20an%20entire%20city%2Dsized

[20] National Instruments. 2020. Peak Signal-to-Noise Ratio as an Image Quality Metric. https://www.ni.com/en-us/innovations/white-papers/11/peak-signal-to-noise-ratio-as-an-image-quality-metric.html

[21] Paras Jain and Vipin Tyagi. 2016. A survey of edge-preserving image denoising methods. *Information Systems Frontiers* 18, 1 (2016), 159–170.

[22] Harpreet Kaur, Deepika Koundal, and Virender Kadyan. 2021. Image fusion techniques: a survey. *Archives of Computational Methods in Engineering* 28, 7 (2021), 4425–4447.

[23] T.S. Kelso. 2021. NORAD Two-Line Element Sets Current Data. Retrieved October 20, 2021 from https://celestrak.com/NORAD/elements/

[24] Gerhard Krieger, Alberto Moreira, Hauke Fiedler, Irena Hajnsek, Marian Werner, Marwan Younis, and Manfred Zink. 2007. TanDEM-X: A satellite formation for high-resolution SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 45, 11 (2007), 3317–3341.

[25] Matan Leibovich, George Papanicolaou, and Chrysoula Tsogka. 2021. Correlation based Imaging for rotating satellites. *SIAM Journal on Imaging Sciences* 14, 1 (2021), 271–303.

[26] Kin K. Leung. 2021. Effects of Noise on FM. Retrieved October 21, 2021 from https://www.commsp.ee.ic.ac.uk/~kkleung/communications2_2009/Lecture5.pdf

[27] Cristoff Martin, Alexander Geurtz, and J Ottersten. 2004. Spectrally efficient mobile satellite real-time broadcast with transmit diversity. In *IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004*, Vol. 6. IEEE, 4079–4083.

[28] Lockheed Martin. 2018. AWS and Lockheed Martin. Retrieved October 20, 2021 from https://news.lockheedmartin.com/news-releases?item=128622

[29] Ussanai Nithirochananont, Michail Antoniou, and Mike Cherniakov. 2020. Passive coherent multistatic SAR using spaceborne illuminators. *IET Radar Sonar Navig* 14 (2020), 628–636.

[30] NOAA. 2009. User's Guide. https://noaasis.noaa.gov/NOAASIS/pubs/Users_Guide-Building_Receive_Stations_March_2009.pdf

[31] NOAA. 2020. STAR Calibration Center. Retrieved October 21, 2021 from https://ncc.nesdis.noaa.gov/VIIRS/VIIRS_Image_Search.php

[32] NOAA. 2021. High Resolution Picture Transmission (HRPT). https://www.noaasis.noaa.gov/POLAR/HRPT/hrpt.html

[33] Daniel Oltrogge, Ramrath AGI, and AGI Jens. 2014. Parametric Characterization of SGP4 Theory and TLE positional accuracy. (2014).

[34] GF Page. 2005. MULTIPLE VIEW GEOMETRY IN COMPUTER VISION, by Richard Hartley and Andrew Zisserman, CUP, Cambridge, UK, 2003, vi+ 560 pp., ISBN 0-521-54051-8.(Paperback£ 44.95). *Robotica* (2005).

[35] Daniele Perissin, Claudio Prati, Marcus E Engdahl, and Yves-Louis Desnos. 2006. Validating the SAR wavenumber shift principle with the ERS–Envisat PS coherent combination. *IEEE Transactions on Geoscience and Remote Sensing* 44, 9 (2006), 2343–2351.

[36] Ioannis Rekleitis, Eric Martin, Guy Rouleau, Régent L'Archevêque, Kourosh Parsa, and Eric Dupuis. 2007. Autonomous capture of a tumbling satellite. *Journal of Field Robotics* 24, 4 (2007), 275–296.

[37] Satnogs.org. 2020. Satnogs Wiki. Retrieved October 20, 2021 from https://wiki.satnogs.org/Main_Page

[38] K Kirk Shung. 2009. High frequency ultrasonic imaging. *Journal of medical ultrasound* 17, 1 (2009), 25–30.

[39] sigidwiki.com. 2020. Low Rate Picture Transmission. https://www.sigidwiki.com/wiki/Low_Rate_Picture_Transmission_(LRPT)

[40] Vaibhav Singh, Akarsh Prabhakara, Diana Zhang, Osman Yağan, and Swarun Kumar. 2021. A community-driven approach to democratize access to satellite ground stations. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 1–14.

[41] DR Sowmya, P Deepa Shenoy, and KR Venugopal. 2017. Remote sensing satellite image processing techniques for image classification: a comprehensive survey. *International Journal of Computer Applications* 161, 11 (2017), 24–37.

[42] M2 Antenna Systems. 2020. LEO-Pack, 436CP16/2MCP8A. Retrieved August 21, 2020 from https://www.m2inc.com/FGLEOPACK

[43] Chunwei Tian, Yong Xu, Lunke Fei, and Ke Yan. 2018. Deep learning for image denoising: A survey. In *International Conference on Genetic and Evolutionary Computing*. Springer, 563–572.

[44] TinyGS. 2020. TINYGS. Retrieved October 20, 2021 from https://tinygs.com/

[45] Peter Van Wie and Maurice Stein. 1977. A landsat digital image rectification system. *IEEE Transactions on Geoscience Electronics* 15, 3 (1977), 130–137.

[46] Deepak Vasisht, Jayanth Shenoy, and Ranveer Chandra. 2021. L2D2: Low latency distributed downlink for LEO satellites. In *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*. 151–164.

[47] Mineo Wakita. 2021. Satellite Frequency List. Retrieved October 10, 2021 from http://www.ne.jp/asahi/hamradio/je9pel/satslist.htm

[48] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.

[49] Satnogs Wiki. 2020. Rotators. https://wiki.satnogs.org/Rotators

[50] Satnogs Wiki. 2021. Ground Station Antennas. Retrieved October 21, 2021 from https://wiki.satnogs.org/Antennas

[51] Feng Yu, Zhen He, Bing Qiao, and Xiaoting Yu. 2014. Stereo-vision-based relative pose estimation for the rendezvous and docking of noncooperative satellites. *Mathematical Problems in Engineering* 2014 (2014).

[52] Bai Zhao, Huaicong Kong, Xiaoyu Liu, Min Lin, Jian Ouyang, and Wei-Ping Zhu. 2021. Transmit diversity and performance analysis for aeronautical broadband satellite communication systems. *Physical Communication* 48 (2021), 101424.

[53] Yufeng Zheng. 2011. *Image fusion and its applications*. BoD–Books on Demand.