

# Resampling Estimation based RPC Metadata Verification in Satellite Imagery

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

24-01-2022 / 28-01-2022

CITATION

Gudavalli, Chandrakanth; Goebel, Michael; Nanjundaswamy, Tejaswi; Nataraj, Lakshmanan; Chandrasekaran, Shivkumar; Manjunath, B.S. (2022): Resampling Estimation based RPC Metadata Verification in Satellite Imagery. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.19026266.v1>

DOI

[10.36227/techrxiv.19026266.v1](https://doi.org/10.36227/techrxiv.19026266.v1)

# Resampling Estimation based RPC Metadata Verification in Satellite Imagery

Chandrakanth Gudavalli<sup>§,1</sup>, Michael Goebel<sup>§,1,2</sup>, Tejaswi Nanjundaswamy<sup>1</sup>, Lakshmanan Nataraj<sup>1</sup>, Shivkumar Chandrasekaran<sup>1,2</sup>, and B.S. Manjunath<sup>1,2</sup>

<sup>1</sup>Mayachitra, Inc.

<sup>2</sup>Electrical and Computer Engineering Department, UC Santa Barbara  
Santa Barbara, California, USA

**Abstract**—Recent advances in machine learning and computer vision have made it simple to manipulate a variety of images, including satellite images. Most of the commercially available satellite images go through the process of orthorectification to remove potential distortions due to terrain variations. This orthorectification process typically involves the use of rational polynomial coefficients (RPC) that geometrically remap the pixels in the original image to the rectified image. This paper proposes a new method to verify the authenticity of these orthorectified images with respect to the associated RPC metadata. The steps include calculating the Residual Discrete Fourier Transform (DFT) pattern from the image using a linear predictor based residual spectral analysis and comparing with expected residual DFT pattern using the RPC metadata associated with the image. If the metadata associated with orthorectified image is the correct one, then both the DFT patterns should have high structural similarity. We use SSIM (Structural Similarity Index Metric) to quantify the similarity and thereby verify if the data has been tampered or not. Detailed experimental results are presented to demonstrate the high accuracy of the proposed method in detecting manipulations.

**Index Terms**—Resampling Estimation, Digital Image Forensics, Metadata Tampering Detection, Signal Processing.

## I. INTRODUCTION

There has been an exponential increase in the number of commercial, public and defense-oriented satellites, and with it concerns over potential manipulation or misuse of such data. Unlike consumer-oriented images, satellite imaging data often has a significant amount of additional metadata, including location, camera orientation, and are typically orthorectified. This orthorectification involves the removal of perspective distortion due to imaging angle, terrain, and lens distortion.

While detecting manipulations in digital images have been well studied [1], [2], [3], there are fewer studies in detecting manipulations in satellite images and even lesser when it comes to tampering of metadata. Satellite images, for the most part, are orthorectified using rational polynomial coefficients (RPC) and Digital Elevation Models (DEM). The RPC coefficients define a best-fit mapping from latitude, longitude, and elevation to pixel coordinates. These coefficients are determined from the camera location, orientation, and parameters intrinsic to the imaging system. Such RPC coefficients are provided in the image metadata. The DEMs contain a dense grid of elevation measures at different points around the globe,

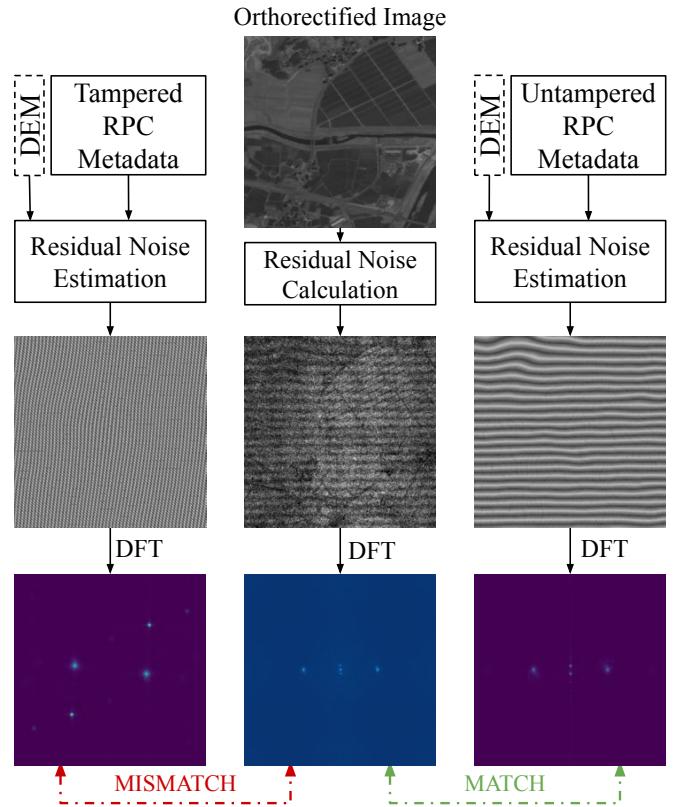


Fig. 1: Visualization of Expected Residual DFT patterns when RPC Metadata is tampered (left) and Untampered (right), with respect to Residual DFT pattern of the image (middle)

and are freely available online. The combination of the RPC coefficients with the DEMs allows for the transformation of the captured image into an orthorectified view.

RPC metadata associated with an image is essential in obtaining the pixel coordinates of a given object or a given lat-long in the unrectified image. Tampering could be used by adversaries to mask the true locations of objects or other geo-features of interest. Tampering of RPC metadata associated with orthorectified satellite imagery raises questions and suspicions on the authenticity of image content.

Our approach to detecting the tampering is based on resampling estimation. Given an image, there exists a noise associated with it. Here noise can be treated as the deviation

<sup>§</sup>Equal contribution

from ideal pinhole camera model. Once image is resampled using an affine transformation, the noise variance fluctuates periodically across the image. DFT of noise variance in the resampled image can be calculated using the method proposed in Kirchner et. al [4], which we refer it as “Residual DFT Pattern” (a sample is shown in Figure 1).

To determine the authenticity of the RPC metadata, we compare the Estimated DFT pattern with an expected pattern, computed using only the RPC and DEM data. Using the RPC + DEM metadata associated with the image, we resample a predefined checkerboard pattern and estimate the DFT of noise variance in the original image. Both the DFTs show structural similarity, as shown in Figure 1, if the metadata used to resample/orthorectify the checkerboard pattern is same as the metadata used to resample the satellite image. Therefore, we use the structural similarity between the two DFT patterns to verify the authenticity of associated RPC metadata.

The main contribution of this paper is the resampling estimation method for robust verification of tampering in orthorectified satellite imagery with respect to the associated RPC metadata. To be specific, we propose technique to calculate *Expected Residual DFT pattern* from the RPC and DEM, which maintains structural similarity with *Residual DFT pattern* extracted only from the image, only when the metadata is not tampered.

The rest of this paper is organized as follows. Section II gives an overview of the related work in image forensics, metadata tampering detection, and resampling estimation. Section III describes the proposed method to verify the authenticity of associated metadata. In Section IV, we provided detailed experimental results to validate the proposed RPC authentication. We conclude the paper with Section V by stating the pros and cons of our technique.

## II. BACKGROUND

There is extensive recent work on image forensics, including techniques to detect image splicing based forgeries ([5], [6]), copy-move forgeries ([7], [8], [1]), image retouching ([9], [10]), seam carving ([11], [12]) and image resampling ([13], [14]). Other common manipulations include machine learning based forgeries, typically manipulated using Generative Adversarial Networks (GANs) ([15], [16]). Deepfake videos ([17], [18]) are the perfect examples of this trend.

The majority of the image forensic techniques are developed to work for images captured using consumer cameras ([19], [20]), which differ in nature from images captured using satellite sensors significantly (e.g., different compression schemes, color channels, camera sensor technologies, radiometric corrections, orthorectification based post processing and so on). Hence, many state-of-the-art forensic techniques are observed to not perform well when applied blindly to satellite images ([21], [22]) as they are not trained to work for satellite imagery.

To address this issue, several forensic tools that work for satellite/overhead imagery have recently been proposed by just using satellite imagery instead of consumer images for training and testing purposes. Yarlagadda et al. [21] presented

a supervised technique that makes use of a conditional GAN to detect and localize splicing forgeries in satellite images by estimating a forgery mask. Horváth [2] proposed a GAN based technique that encodes patches from a test image into a low dimensional vector space which is then used by a support vector machine (SVM) and detect the presence of forgeries at patch level. Bartusiak et al. [22] proposed a kernel-based one-class classification method, named Sat-SVDD, to detect splicing forgeries with the help of support vector data descriptors (SVDD). Luqman et al. [23] proposed a machine learning based passive method to predict the regions of satellite images doctored using inpainting technique.

While several data dependent image forgery detection models are proposed to verify the authenticity of overhead satellite imagery, verifying the authenticity of satellite imagery by taking advantage of metadata associated with these images haven't been explored thoroughly. Specifically, there are no techniques to verify the authenticity of RPC metadata associated with the orthorectified satellite imagery.

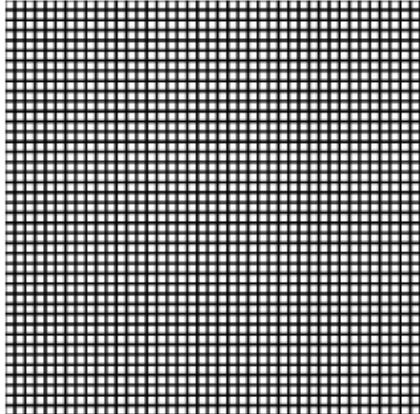
Since the process of orthorectification creates a new, warped set of image pixel locations, resampling must be used to produce the orthorectified image. Altering of the RPC coefficients will affect the warping pattern used, and therefore the resampling. There are several methods used for resampling detection and/or estimation ([24], [25], [4], [26], [3], [13]). In our proposed approach, we selected a fixed linear predictor based residual spectral analysis as described in Kirchner et. al [4]. This method offers fast prediction for large images, which makes the technique reliable for satellite images as they tend to have larger dimensions. Images that we worked on are typically of size 20,000 x 8,000. This method also calculates relatively unique features for a variety of scaling, rotation, and sheer factors.

### A. Orthorectification and RPC Metadata

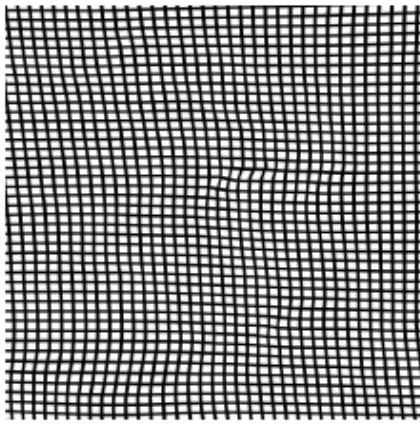
Orthorectification is the process of transforming an image onto its upright planimetry map by removing the perspective angle. Orthorectification is done using Rational Polynomial Coefficients (RPCs) based on empirical models that relate the geographic location (latitude/longitude, denoted by  $X, Y$ ) and the surface elevation data (denoted by  $Z$ ) to map the row and column positions (denoted by  $r, c$ ) through two rational polynomials [27]. Satellite sensor models are empirical mathematical models that relate image coordinates (row and column position) to latitude and longitude using the terrain surface elevation. The name Rational Polynomial derives from the fact that the model is expressed as the ratio of two cubic polynomials. A single image involves two rational polynomials, one each for computing row and column position as shown in Eqn 1.

$$r = \frac{P_1(X, Y, Z)}{P_2(X, Y, Z)}, \quad c = \frac{P_3(X, Y, Z)}{P_4(X, Y, Z)} \quad (1)$$

where,  $P_1, P_2, P_3$ , and  $P_4$  are cubic polynomials, each with 20 coefficients (which are referred as RPC Metadata) as shown in Eqn 2.



(a) Synthetic Grid (Before Orthorectification)



(b) Synthetic Grid (After Orthorectification)

Fig. 2: Effect of Orthorectification on a grid of vertical and horizontal lines.

$$\begin{aligned}
 P_*(X, Y, Z) = & a_{*_1} + a_{*_2}X + a_{*_3}Y + a_{*_4}Z + a_{*_5}X^2 + \\
 & a_{*_6}XY + a_{*_7}XZ + a_{*_8}Y^2 + a_{*_9}YZ + \\
 & a_{*_10}Z^2 + a_{*_11}X^3 + a_{*_12}X^2Y + a_{*_13}X^2Z + \\
 & a_{*_14}Y^3 + a_{*_15}XY^2 + a_{*_16}Y^2Z + a_{*_17}XYZ + \\
 & a_{*_18}XZ^2 + a_{*_19}YZ^2 + a_{*_20}Z^3
 \end{aligned} \quad (2)$$

where, \* belongs to 1, 2, 3, or 4.

The coefficients of these two rational polynomials (shown in Eqn 1) are computed as the best fit mapping from spatial location (X, Y, Z) to pixel coordinates (r, c). This is done by considering the camera's orbital position, orientation, and corresponding physical sensor model.

Using the unrectified satellite image, its RPC Metadata, and a Digital Elevation Map (DEM) to provide the elevation values, an unrectified image is resampled to generate an orthorectified image. Figure 2 shows a visualization of this transformation, by passing a grid of vertical and horizontal lines through the warping function. DEMs with 30-meter resolution produced by the United States Geological Survey (USGS) are available for free download for any area in the United States, and 10-meter USGS DEMs are available in

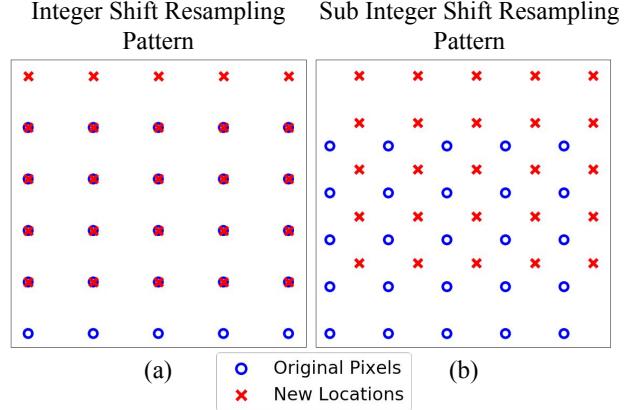


Fig. 3: Some example resampling patterns. Only the sub-integer shifts impact the noise variance patterns. (a) Pattern where noise variance is  $\sigma^2$ . (b) Pattern where noise variance is  $0.25\sigma^2$ .

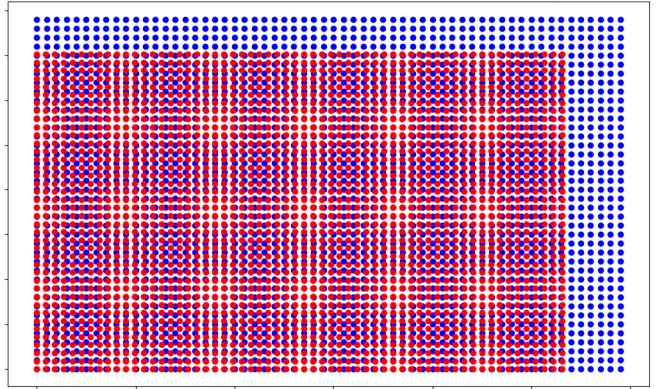


Fig. 4: Resampling spatial mapping for a scaling example (Scale factor of 0.9; blue: original locations; red: new locations). These sampling patterns move in and out of phase with each other to form a periodic pattern. This pattern is reflected in the noise variance of the resampled image.

most areas of the globe. Since resampling is a key step in the process of generating an orthorectified image, detecting inconsistencies caused due to resampling can be used to quantify the authenticity of orthorectified image with respect to RPC metadata.

#### B. Calculation of Residual DFT Pattern

While there are several methods for resampling detection and/or estimation, we selected a fixed linear predictor based residual spectral analysis as described in Kirchner et. al [4]. This method offers faster prediction, which is essential for satellite images with sizes close to 20000 x 8000 pixels. This method first estimates an image noise signal by applying a fixed linear filter. Here noise refers to any deviation in the sample from an ideal pinhole camera model. It is shown that different resampling patterns will create unique, periodic artifacts in the noise variance, and can be analyzed through the DFT of noise variance, which we refer as Residual DFT

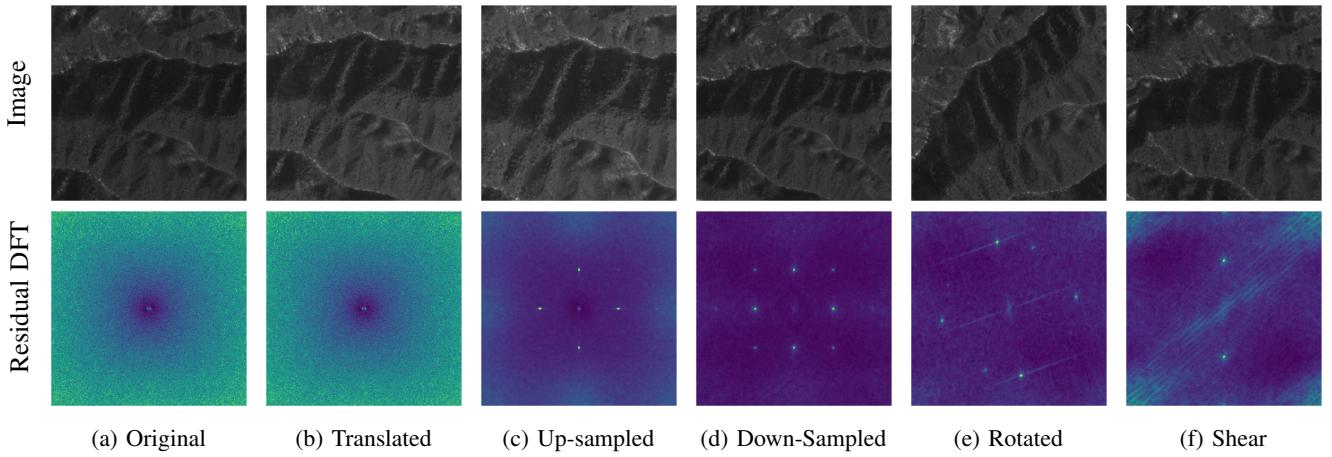


Fig. 5: Some toy examples showing the image residual DFT (2nd row) under different transformations.

Pattern. We now briefly describe this resampling estimation method as it is used to calculate the residual DFT pattern.

Kirchner et. al [4] assumes that the pixel noise in the image captured by the camera is with zero mean and constant variance ( $\sigma^2$ ). Resampling will cause the noise variance to fluctuate depending on the location. For the points where resampled pixels map directly onto one of the positions of original pixels, the variance will also be  $\sigma^2$ . Points that lie equidistant from its 4 nearest neighbors will have a noise variance of only  $0.25\sigma^2$ . Visual representations of both are shown in Figure 3. All other points in the resampling pattern will have a noise variance which lies between these two extremes.

It is shown by [4] that affine resampling methods introduce periodic patterns in the pixel noise variance present in the resampled image. An example of this is shown in Figure 4. This figure represents the pixel displacements that occurred due to resampling techniques in which the image is scaled by a factor of 0.9. The new pixel locations will be coming in and out of phase with the original pixel locations, causing periodic patterns in the noise variance of resampled image. Given this model of the resampling process, we only need a method to estimate the pixel noise variance. Then, unique periodic patterns will be visible in the DFT of the variance estimate, which is referred as Residual DFT pattern.

To estimate the noise variance in an image the following procedure is used by Kirchner et. al [4]. First, a high pass filter is used to remove a sufficient amount of image content. The following convolution kernel is used for the tests:

$$\begin{bmatrix} -1/4 & 1/2 & -1/4 \\ 1/2 & -1 & 1/2 \\ -1/4 & 1/2 & -1/4 \end{bmatrix}$$

To estimate the noise value at each pixel, the method applies the above linear, high pass filter to remove the image content. This filtered image (denoted by  $e$ ) is treated as an estimate of the noise values at each pixel. The method estimates the noise variance (denoted by  $p$ ) similar to Popescu

and Farid's Gaussian distribution based calculation [28] as shown in Eqn 3.

$$p = \lambda \exp\left(-\frac{|e|^\tau}{\sigma}\right) \quad (3)$$

where  $\lambda, \sigma > 0$  and  $\tau \geq 1$  are controlling parameters. The controlling parameters have been fixed by the method to  $\lambda = 1, \sigma = 1$  and  $\tau = 2$ . The DFT of the estimated noise variance ( $p$ ) is referred to as *Residual DFT pattern*. Some toy examples showing the image residual DFT patterns under different transformations are shown in Figure 5.

A distinction from previous works in resampling detection is that the resampling pattern for these satellite images is not affine. While an affine transformation will produce periodic artifacts, and discrete points in the Expected DFT spectrum, the RPC+DEM resampling patterns will not. However, these patterns are very close to being locally affine for small patches, and will instead form a cloud of points in the DFT pattern.

### III. PROPOSED METHOD

#### A. Calculation of Expected Residual DFT Pattern

We calculate the DFT of “expected noise variance” at each point in the orthorectified image, which is referred to as Expected Residual DFT Pattern.

The expected noise variance is estimated at each point in the orthorectified image by using the RPC metadata associated with the image. In order to estimate the variance at a given point, we exploit the fact that variance is inversely proportional to distance between the new orthorectified pixel location and its nearest neighbor in the unrectified image. So, we calculate L1 distance between them (Several distance types were tested experimentally, and L1 distance was chosen). We compute the DFT of the calculated distance, which is same as Expected DFT Pattern as the distance is inversely proportional to the variance.

We get these distances using only the forward warp function by using the following procedure. An array,  $Y$ , of the same height and width as the image (which is referred to as synthetic grid) is initialized with 4 channels. Then, it is filled

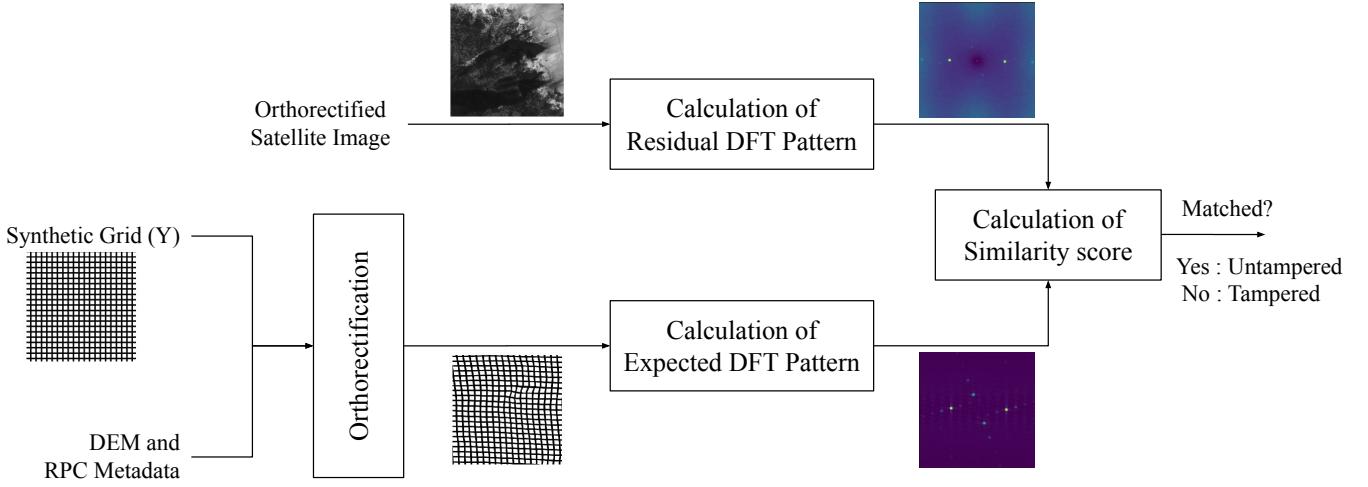


Fig. 6: Overview of proposed framework to detect tampering in RPC Metadata.

with values such that any  $2 \times 2$  block contains 4 orthogonal vectors as shown in Eq 4.

$$Y = \begin{bmatrix} a & b & a & b & \dots \\ c & d & c & d & \\ a & b & a & b & \\ c & d & c & d & \\ \vdots & & & & \ddots \end{bmatrix} \quad (4)$$

where,  $a = [255, 0, 0, 0]$ ,  $b = [0, 255, 0, 0]$ ,  $c = [0, 0, 255, 0]$ ,  $d = [0, 0, 0, 255]$ .

This matrix is passed through the same transformation pipeline as the image. Assuming that the transformation can be approximated to a bilinear interpolation, from the transformed version of  $Y$ , the bilinear interpolation formula can then be reversed (as described in Appendix A) to calculate the distance matrix which represents the distance between each of the new orthorectified pixel location and its nearest neighbor in the unrectified image. This ensures that, it is required to provide only the forward warping function.

We calculate the DFT magnitude of this distance matrix, which we refer to as the Expected DFT pattern. This will later be compared with the residual DFT pattern to measure the structural similarity between the both. Before comparing both the DFT patterns, we high pass filter the residual DFT pattern through multiplication by a cone,  $C$  (shown in Eqn 5). We do it to suppress the strong and less informative low frequency components, and level out the noise floor in it.

$$C(row, col) = \sqrt[4]{(2 * row - h)^2 + (2 * col - w)^2} \quad (5)$$

where  $h, w$  are height and width of DFT matrix.

This entire process is summarized into a flow chart that is shown in Figure 6. In this section, all processes in the flow chart are defined except for computing the mismatch between two DFT patterns. Section III-B investigates different methods of computing the mismatch score.

#### B. Similarity score calculation between DFT Patterns

Given two DFT patterns, a metric to quantify the similarity between them is required. One of the options would be to use Mean squared error (MSE) as a metric to quantify the dissimilarity between two DFT arrays (say  $x, y$ ) as shown in Eqn 6.

$$d_{MSE} = \frac{1}{M * N} \sum_{i=1}^N \sum_{j=1}^M (x_{i,j} - y_{i,j})^2 \quad (6)$$

where  $M, N$  are number of rows and number of columns of the DFT arrays respectively.

Otherwise, both the two-dimensional arrays can be flattened into a single dimension and compute the cosine similarity score between them as shown in Eqn 7.

$$\text{cosine\_similarity} = \frac{\alpha \cdot \beta}{\|\alpha\| \cdot \|\beta\|} \quad (7)$$

where  $\alpha, \beta$  are the flattened DFT arrays.

Eqn 6 and Eqn 7 show that both MSE and cosine similarity metric take every pixel into account and perform a one-to-one comparison when computing the similarity score.

Since, we observed from our experiments that, the intensity and position of peaks in DFT patterns can vary slightly around a specific area, image similarity measures that capture “structural” similarity are of interest. We used Structural Similarity Index Metric (SSIM) [29] to calculate the similarity score between two DFT arrays. SSIM score between two discrete signals is calculated as follows. Let  $x, y$  be the two DFT arrays that correspond to a given image patch. Let  $\mu_x, \sigma_x^2$  and  $\sigma_{xy}$  be the mean of  $x$ , variance of  $x$ , and the covariance of  $x$  and  $y$ , respectively. Approximately,  $\mu_x$  and  $\sigma_x$  can be viewed as estimates of the luminance and contrast of  $x$ , and  $\sigma_{xy}$  measures the the tendency of  $x$  and  $y$  to vary together, thus an indication of structural similarity. SSIM compares luminance ( $l$ ), contrast ( $c$ ), and structure ( $s$ ) of  $x, y$

(using Eqn 8, Eqn 9, and Eqn 10, respectively) and the overall similarity score is computed using Eqn 13.

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (8)$$

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (9)$$

$$s(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (10)$$

$$C_3 = C_2/2 \quad (11)$$

where  $C_1$ ,  $C_2$  and  $C_3$  are small constants given by

$$C_1 = (K_1 L)^2, C_2 = (K_2 L)^2 \text{ and } C_3 = C_2/2, \quad (12)$$

$L$  is the dynamic range of the pixel values ( $L = 255$  for 8 bits/pixel gray scale images), and  $K_1 \ll 1$  and  $K_2 \ll 1$  are two scalar constants. The general form of SSIM between signal  $\mathbf{x}$  and  $\mathbf{y}$  is defined as:

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})^\alpha \cdot c(\mathbf{x}, \mathbf{y})^\beta \cdot s(\mathbf{x}, \mathbf{y})^\gamma] \quad (13)$$

where  $\alpha$ ,  $\beta$  and  $\gamma$  are parameters to define the relative importance of the three components. We use  $\alpha, \beta, \gamma = 1$  to give equal importance to luminance ( $l$ ), contrast ( $c$ ), and structure ( $s$ ). Hence the resulting SSIM index is given by

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (14)$$

Since SSIM measures similarity (i.e., 1 implies matched, and 0 implies mismatched), we use  $(1 - \text{SSIM}(\mathbf{x}, \mathbf{y}))$  as the metric to calculate the distance (or dissimilarity) between the two DFT patterns in our experiments. Figure 7 and Figure 8 show sample scenarios explaining the reason for choosing SSIM over MSE and cosine similarity metrics respectively.

SSIM based scores were concentrated in a very small range. So we normalized the scores using a sigmoid function,  $f(x)$  in Eq 15, to be well spread between 0 (matched) and 1 (mismatched), thereby easily differentiate matched and mismatched pairs.

$$f(x) = \frac{1}{1 + e^{-\lambda_e(x - \mu_e)}} \quad (15)$$

We empirically estimated  $\lambda_e, \mu_e$  (which determine the shape of the sigmoid curve) from the outputs of experiments on larger datasets. Details regarding the datasets and the experiments that are carried out on them are discussed in Section IV.

Given a large satellite image and associated RPC metadata, we divide the image into non-overlapping patches and compute both the DFT patterns for each patch. We calculate the SSIM score for each patch and generate a heatmap of SSIM scores for the entire image. Median of patch-wise SSIM scores is treated as the overall tampering score of the image, whereas heatmap can be used to determine where and how the match and mismatch differ.

For example, heatmaps for a given pair of images are shown in Figure 9. We can now clearly see the smaller (blue) score values for the matched case (top-left and bottom-right images) and larger (red) score values for the mismatched case (top-right and bottom-left).

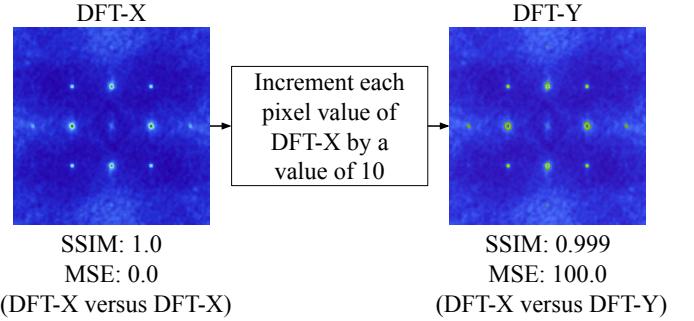


Fig. 7: Sample demonstration to show that SSIM is more apt than MSE for our use case. Even though both the DFT patterns look alike, MSE is unable to capture the similarity unlike SSIM.

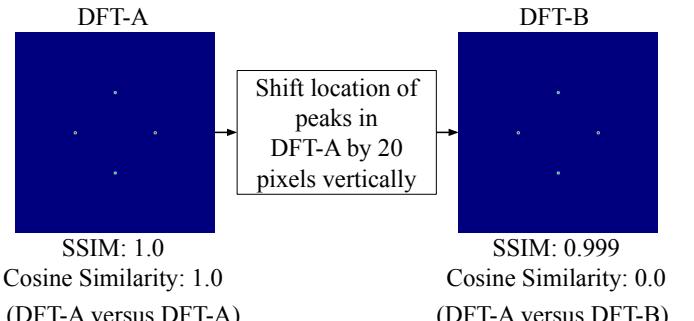


Fig. 8: Sample demonstration to show that SSIM is more apt than Cosine Similarity for our use case. Even though both the DFT patterns look alike, cosine similarity is unable to capture the similarity unlike SSIM.

#### IV. EXPERIMENTS

This section details the experiments that are carried out to verify the authenticity of satellite images using proposed technique.

##### A. Dataset

Level 1B data (in GeoTIFF format, with RPC Metadata associated) from Orbview-3 satellite [30] is collected from United States Geological Survey (USGS) Earth Explorer [31]. For elevation maps that are required for orthorectifying the images, we used Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global data [32].

Samples from different regions of the globe with both flat and hilly terrains are used to carry out the experiments. Table I shows the number of samples collected from each region. Each sample in the dataset has pixel information and corresponding metadata associated with it. We created a dataset of tampered and untampered samples by replacing the metadata of each

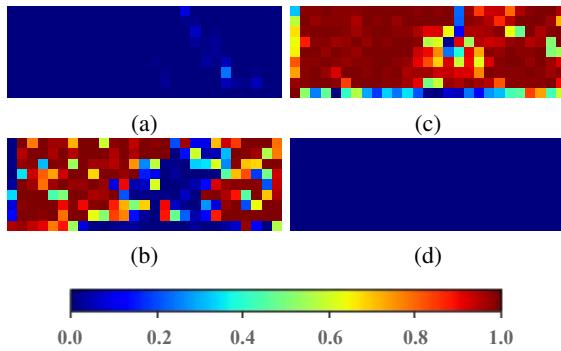


Fig. 9: Sample heatmaps generated using patch wise score. Here patches of size 1024x1024 are used. We compute the patch wise scores under two criteria: (1) RPC coefficients are not exchanged (matched), (2) RPC coefficients of one image is exchanged with the other (mismatch). First image (a) and fourth image (d) show the heatmaps when the RPC coefficients are not exchanged. Second image (c) and third image (d) show the heatmaps when the RPC coefficients are exchanged. We can clearly see that the score values are less when the coefficients are not exchanged.

sample with metadata associated with one of the other samples in dataset.

### B. Experimental results on Japan Dataset

Experiments are carried out on Japan dataset by considering the collected image-metadata pairs as samples of untampered dataset. Tampered dataset is generated by randomly exchanging the metadata between samples. We calculated the expected DFT pattern and residual DFT pattern of each sample as described in Section III-A and Section II-B respectively, by dividing images into non-overlapping patches of size 1024 x 1024. As described in Section III-B, normalized SSIM score between DFT patterns of each patch is calculated. Median of scores of all patches in a given image is considered as the overall tampering score of the image and this score is used as the key to detect tampering in metadata. Binary classification of these scores resulted in:

- Area under ROC curve (AUC) of 0.9969
- Maximum accuracy of **99.15%**
- Tampered detection accuracy of 99.65% (Percentage of tampered samples that are detected as tampered)

Region	Number of Images
Japan (Hilly Terrain)	1708
Northern Europe (Estonia, Latvia, Lithuania) (Flat Terrain)	88
South America (Paraguay) (Flat Terrain)	848
West Africa (Western Sahara region) (Flat Terrain)	355

TABLE I: Data collected from different regions of the globe to test the proposed method.

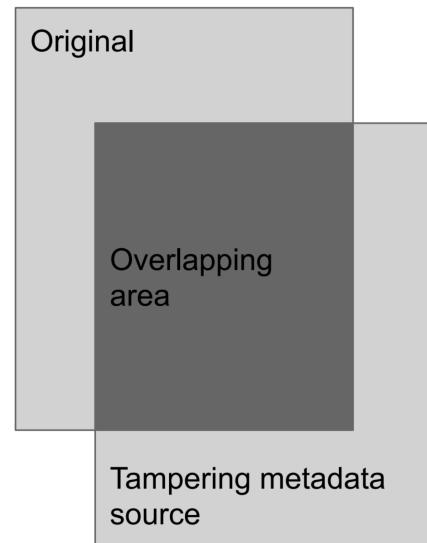


Fig. 10: Illustration of overlapping satellite images.

- Untampered detection accuracy of 98.65% (Percentage of untampered samples that are detected as untampered)

We repeated the above experiment for various patch sizes and the corresponding results are shown in Table II. As we got highest accuracy for patch size of 1024x1024, we finalized on breaking the images into 1024x1024 patches.

Patch size	AUC	Accuracy (%)
2048x2048	0.9978	98.92
1024x1024	0.9969	<b>99.15</b>
512x512	0.9951	99.00
256x256	0.9956	98.68
128x128	0.9958	98.71

TABLE II: Performance for varying patch size

### C. Experiments by temporally exchanging metadata

Experiments presented in Section IV-B are carried out by randomly exchanging metadata between different samples in a dataset. But, a more purposeful manipulation is exchanging metadata between samples corresponding to the same GPS coordinates that are captured at different timestamps. To recreate this, for each image, we search for a maximum overlapping pair captured at different timestamp in the dataset. We used the image pairs with overlap percentage (overlap area/original area, as shown in Figure 10) above a threshold to exchange metadata between them.

An example pair of Orbview-3 images with overlap area threshold greater than 85% is shown in Figure 11. With 85% overlap area threshold, only 164 images (and 328 samples) remain from the 1708 images of Japan region. Conducting tampered sample detection experiments with patch size of 1024x1024 on this small dataset resulted in:

- Area under ROC curve (AUC) of 0.9848
- Maximum accuracy of **97.85%**
- Tampered detection accuracy of 98.16%
- Untampered detection accuracy of 97.55%

Region	Overlap Area Threshold	Number of Samples	AUC-ROC	Maximum Accuracy	Tampered detection Accuracy	Untampered detection accuracy
Japan	0% (Random)	3416 (1708 images)	0.9969	99.15	99.65	98.65
	85% Overlap	328 (164 images)	0.9848	97.85	98.16	97.55
	98% Overlap	94 (47 images)	0.9941	96.81	100.00	93.62
Northern Europe	0% (Random)	174 (87 images)	0.9877	98.27	97.70	98.85
	85% Overlap	138 (69 images)	1.00	100.00	100.00	100.00
	98% Overlap	90 (45 images)	1.00	100.00	100.00	100.00
South America	0% (Random)	1696 (848 images)	0.9991	98.93	98.82	99.05
	85% Overlap	518 (259 images)	0.9891	96.14	95.37	96.91
	98% Overlap	246 (123 images)	0.9893	96.69	96.69	96.69
West Africa	0% (Random)	710 (355 images)	0.9998	99.43	98.87	100.00
	85% Overlap	134 (67 images)	0.9973	97.76	95.52	100.00
	98% Overlap	52 (26 images)	1.00	100.00	100.00	100.00

TABLE III: Experimental results on the datasets from different parts of globe



Fig. 11: Sample pair of images with overlap area threshold greater than 85%.

With 98% overlap area threshold, only 47 images (and 94 samples) remain from the big dataset of 1708 images. Classification results on this dataset resulted in:

- Area under ROC curve (AUC) of 0.9941
- Maximum accuracy of **96.81%**
- Tampered detection accuracy of 100%
- Untampered detection accuracy of 93.62%

For dataset from Japan region, results for purposeful temporal metadata exchange are slightly worse than random metadata exchange. But, this is not the case with the datasets from other regions.

#### D. Testing on flat regions

Orbview-3 satellite data is collected from other regions of the globe with flat regions (< 500 feet variation). Tampering detection experiments are conducted for Northern Europe,

South America and West Africa datasets and corresponding results are described in Table III

The comprehensive results for three different flat regions of the world compared to the Japan dataset, suggest that the performance of our proposed algorithm does not suffer even when there are no significant terrain features that can add more features to the DFT pattern.

#### E. Visual Examples

Few examples visualizing the residual DFT patterns and expected DFT patterns where our method was successful at detecting if the metadata is tampered and untampered are shown in Figure 12 and Figure 13 respectively. Even though our method has shown higher accuracy for RPC tampering detection, we show the rare examples where our method happen to fail in Figures 14, 15.

## V. CONCLUSION

This paper proposes a novel approach to verify the authenticity of orthorectified satellite images using the associated RPC metadata that is used to generate the orthorectified image. We calculate the expected DFT pattern from metadata and residual DFT pattern from pixel content of the image. These two DFT patterns tend to have structural similarity if the RPC metadata is untampered. We use sigmoid normalized SSIM score to classify tampered and untampered samples. Extensive experimental results are provided to demonstrate the effectiveness of the approach to detecting the RPC metadata manipulations.

## APPENDIX A CALCULATION OF EXPECTED DFT

This section describes math behind the calculation of expected DFT patterns from the synthetic input,  $\underline{Y}$ , as defined in Eqn 4.

Let  $\underline{Y}$  be a  $2 \times 2$  sub array from  $\underline{Y}$ . This implies  $\underline{Y}$  can be one of the following:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ (or)} \begin{bmatrix} b & a \\ d & c \end{bmatrix} \text{ (or)} \begin{bmatrix} c & d \\ a & b \end{bmatrix} \text{ (or)} \begin{bmatrix} d & c \\ b & a \end{bmatrix} \quad (16)$$

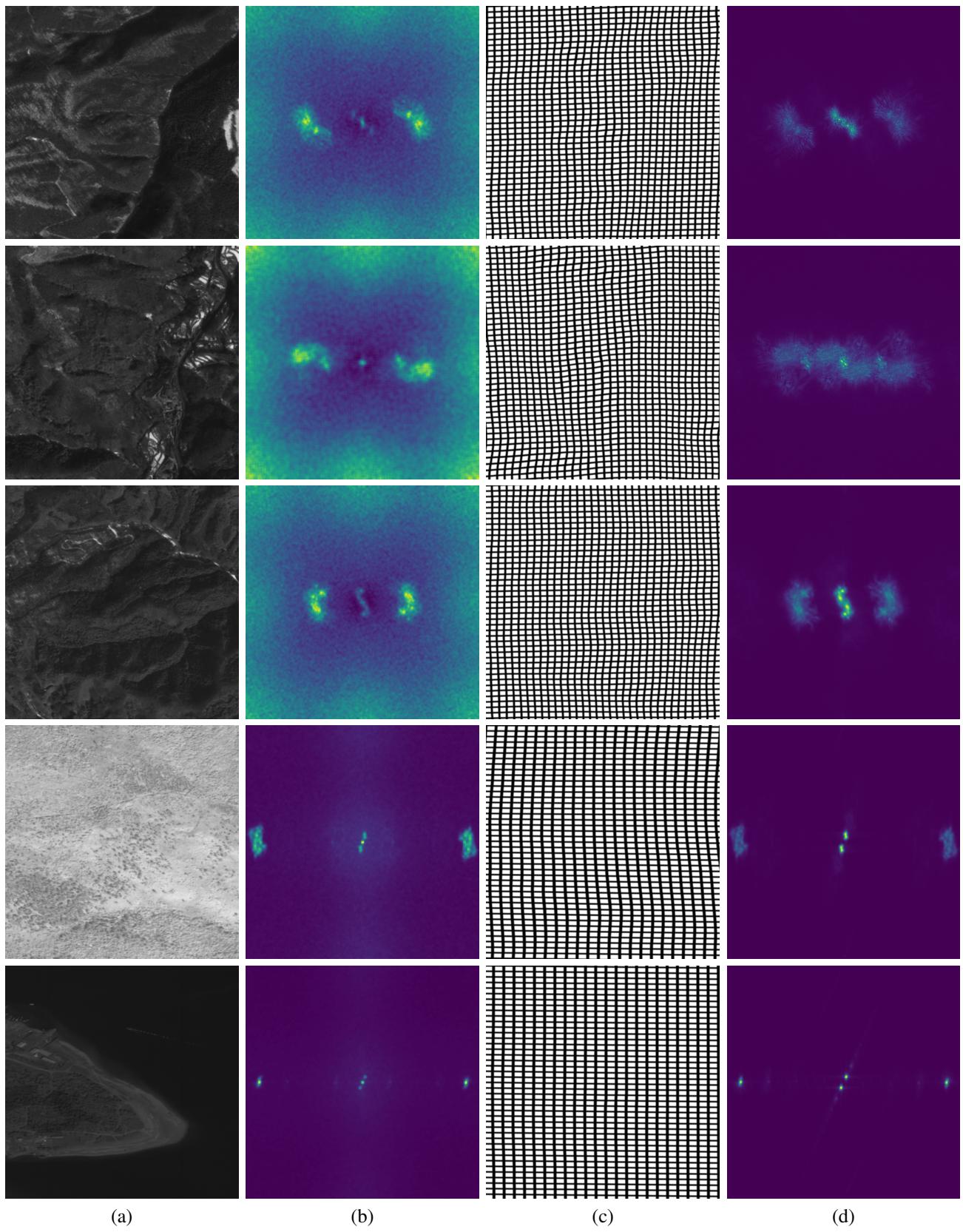


Fig. 12: Sample DFT patterns when RPC metadata is untampered. From left to right: (a) Orthorectified satellite image; (b) Residual DFT Pattern; (c) Rectangular grid of horizontal and vertical lines; (d) Expected DFT Pattern.

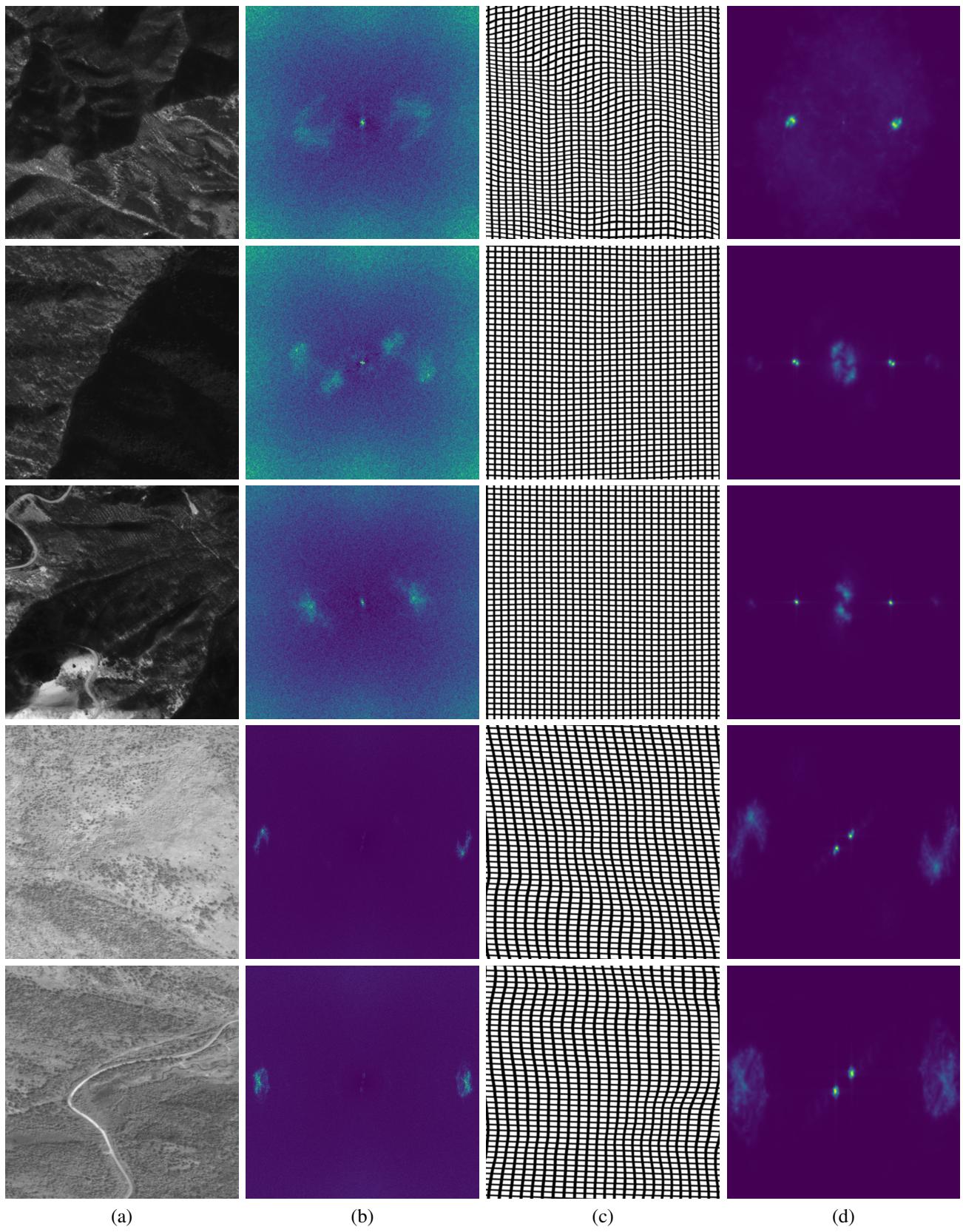


Fig. 13: Sample DFT patterns when RPC metadata is tampered. From left to right: (a) Orthorectified satellite image; (b) Residual DFT Pattern; (c) Rectangular grid of horizontal and vertical lines; (d) Expected DFT Pattern.

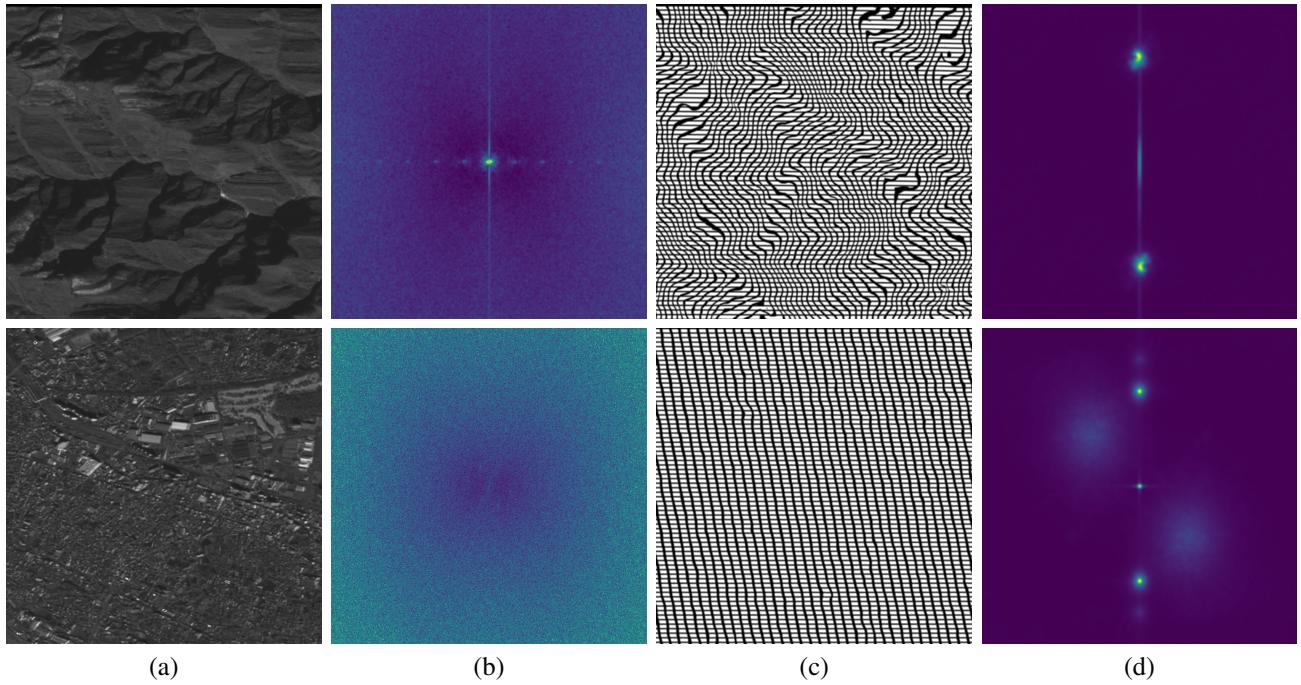


Fig. 14: Sample images where RPC metadata is untampered, but DFT patterns do not have structural similarity. The first example shows more significant distortion in the resampling grid than other examples. The second appears to have a typical resampling pattern, but the image may have been subjected to additional post-processing: (a) Orthorectified satellite image; (b) Residual DFT Pattern; (c) Rectangular grid of horizontal and vertical lines; (d) Expected DFT Pattern.

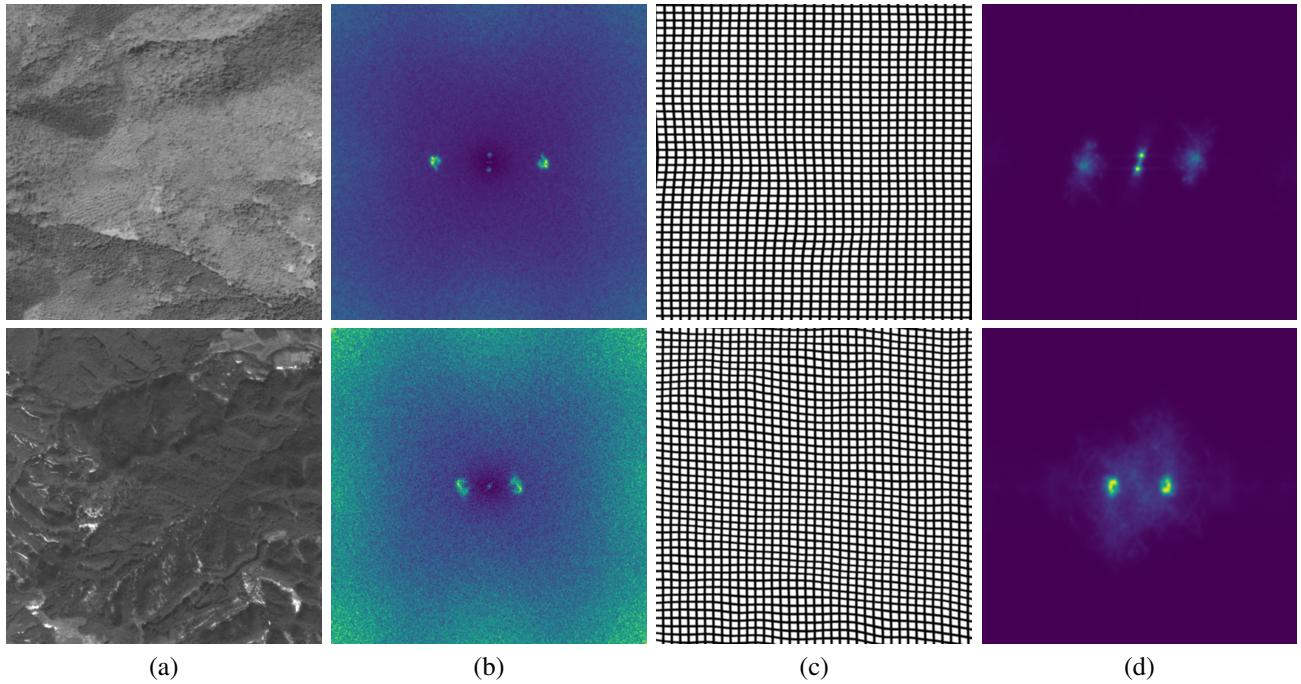


Fig. 15: Sample images where RPC metadata is tampered, but DFT patterns have structural similarity. Our method of tampering randomly selects a new set of RPC coefficients, and there is a small chance that the transformation corresponding to the randomly selected coefficients is almost equal to the original: (a) Orthorectified satellite image; (b) Residual DFT Pattern; (c) Rectangular grid of horizontal and vertical lines; (d) Expected DFT Pattern.

For the illustration purpose, we consider  $\underline{Y}$  as

$$\underline{Y} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad (17)$$

Let 'X' be the interpolated pixel in the transformed image that is generated by using "bilinear interpolation" of the four pixels of  $\underline{Y}$  shown in Eqn 17. Therefore, for some  $0 \leq x, y \leq 1$ , we can say

$$X = [(1-y) \quad y] [\underline{Y}] \begin{bmatrix} 1-x \\ x \end{bmatrix} \quad (18)$$

From Eqn 17 and Eqn 18, we can say that:

$$X = 255 \begin{bmatrix} (1-y)(1-x) \\ (1-y)x \\ y(1-x) \\ yx \end{bmatrix} \quad (19)$$

Let the four channel pixel 'X' be represented using  $X_0, X_1, X_2, X_3$ . This implies,

$$\begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \end{bmatrix} = 255 \begin{bmatrix} (1-y)(1-x) \\ (1-y)x \\ y(1-x) \\ yx \end{bmatrix} \quad (20)$$

From Eqn 20, we can say that:

$$x = \frac{X_1}{X_1 + X_0} \quad (21)$$

$$y = \frac{X_2}{X_2 + X_0} \quad (22)$$

$\underline{Y}$  will not be always as shown in Eqn 17. It can be any one of the four 2x2 arrays shown in 16.

Now, let us assume that:

$$\underline{Y} = \begin{bmatrix} d & c \\ b & a \end{bmatrix} \quad (23)$$

Then, we get:

$$X = 255 \begin{bmatrix} xy \\ y(1-x) \\ (1-y)x \\ (1-x)(1-y) \end{bmatrix} \quad (24)$$

Eqn 24 implies that:

$$\frac{X_1}{X_1 + X_0} = 1 - x \quad (25)$$

$$\frac{X_2}{X_2 + X_0} = 1 - y \quad (26)$$

Eqn 25 and Eqn 26 implies that we get point B in Figure 16 instead of point A. But, distance  $d$  to the "nearest neighbor in unrectified domain" is same in both the cases, which can be calculated as:

$$d = |0.5 - |x - 0.5|| + |0.5 - |y - 0.5|| \quad (27)$$

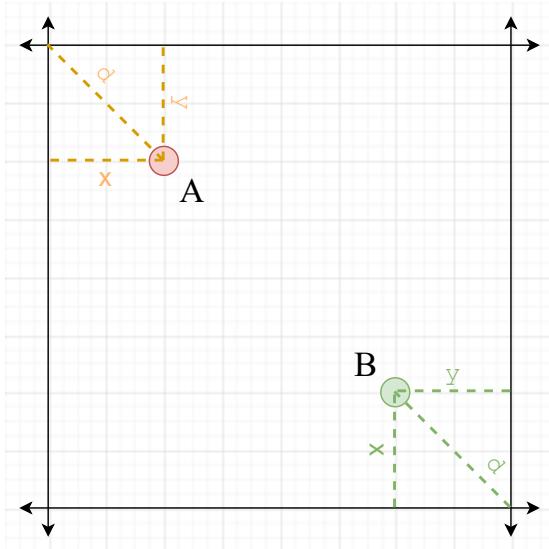


Fig. 16: Resampling spatial mapping for a scaling example

Expected DFT pattern is therefore calculated by computing the 2d DFT of this expected prediction residue ( $d$ ) for all the pixels in rectified domain, which is found to have structural similarity with the residual DFT pattern calculated using the rectified satellite image.

#### ACKNOWLEDGMENT

This work was supported by the National Geospatial-Intelligence Agency (NGA) under Contract No. HM047619C0056. The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the NGA or the U.S. Government.

#### REFERENCES

- [1] J. Li, X. Li, B. Yang, and X. Sun, "Segmentation-based image copy-move forgery detection scheme," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 3, pp. 507–518, 2015.
- [2] J. Horváth, D. Güera, S. K. Yarlagadda, P. Bestagini, F. M. Zhu, S. Tubaro, and E. J. Delp, "Anomaly-based manipulation detection in satellite images," *networks*, vol. 29, p. 21, 2019.
- [3] J. Bunk, J. H. Bappy, T. M. Mohammed, L. Nataraj, A. Flenner, B. Manjunath, S. Chandrasekaran, A. K. Roy-Chowdhury, and L. Peterson, "Detection and localization of image forgeries using resampling features and deep learning," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2017, pp. 1881–1889.
- [4] M. Kirchner, "Fast and reliable resampling detection by spectral analysis of fixed linear predictor residue," in *Proceedings of the 10th ACM Workshop on Multimedia and Security*, ser. MM&Sec 2008. New York, NY, USA: Association for Computing Machinery, 2008, p. 11–20. [Online]. Available: <https://doi.org/10.1145/1411328.1411333>
- [5] F. Hakimi, M. Hariri, and F. GharehBaghi, "Image splicing forgery detection using local binary pattern and discrete wavelet transform," in *2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI)*. IEEE, 2015, pp. 1074–1077.
- [6] M.-J. Kwon, I.-J. Yu, S.-H. Nam, and H.-K. Lee, "Cat-net: Compression artifact tracing network for detection and localization of image splicing," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 375–384.
- [7] A. Islam, C. Long, A. Basharat, and A. Hoogs, "Doa-gan: Dual-order attentive generative adversarial network for image copy-move forgery detection and localization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 4676–4685.

- [8] Y. Li and J. Zhou, "Fast and effective image copy-move forgery detection via hierarchical feature point matching," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1307–1322, 2018.
- [9] C. Rathgeb, C.-I. Satnoianu, N. E. Haryanto, K. Bernardo, and C. Busch, "Differential detection of facial retouching: A multi-biometric approach," *IEEE Access*, vol. 8, pp. 106 373–106 385, 2020.
- [10] Y. Guo, X. Cao, W. Zhang, and R. Wang, "Fake colorized image detection," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 8, pp. 1932–1944, 2018.
- [11] L. Nataraj, C. Gudavalli, T. M. Mohammed, S. Chandrasekaran, and B. Manjunath, "Seam carving detection and localization using two-stage deep neural networks," in *Machine Learning, Deep Learning and Computational Intelligence for Wireless Communication*. Springer, 2021, pp. 381–394.
- [12] A. Sarkar, L. Nataraj, and B. S. Manjunath, "Detection of seam carving and localization of seam insertions in digital images," in *Proceedings of the 11th ACM workshop on Multimedia and security*, 2009, pp. 107–116.
- [13] A. Flennner, L. Peterson, J. Bunk, T. M. Mohammed, L. Nataraj, and B. S. Manjunath, "Resampling forgery detection using deep learning and a-contrario analysis," *Electronic Imaging*, vol. 2018, no. 7, pp. 1–7, 2018.
- [14] T. Qiao, A. Zhu, and F. Retraint, "Exposing image resampling forgery by using linear parametric model," *Multimedia Tools and Applications*, vol. 77, no. 2, pp. 1501–1523, 2018.
- [15] L. Verdoliva, "Media forensics and deepfakes: An overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 910–932, 2020.
- [16] L. Nataraj, T. M. Mohammed, B. Manjunath, S. Chandrasekaran, A. Flennner, J. H. Bappy, and A. K. Roy-Chowdhury, "Detecting gan generated fake images using co-occurrence matrices," *Electronic Imaging*, vol. 2019, no. 5, pp. 532–1, 2019.
- [17] Y. Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019.
- [18] S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano, and H. Li, "Protecting world leaders against deep fakes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.
- [19] L. Bondi, L. Baroffio, D. Güera, P. Bestagini, E. J. Delp, and S. Tubaro, "First steps toward camera model identification with convolutional neural networks," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 259–263, 2016.
- [20] L. Bondi, S. Lameri, D. Güera, P. Bestagini, E. J. Delp, and S. Tubaro, "Tampering detection and localization through clustering of camera-based cnn features," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 1855–1864.
- [21] S. K. Yarlagadda, D. Güera, P. Bestagini, F. Maggie Zhu, S. Tubaro, and E. J. Delp, "Satellite image forgery detection and localization using gan and one-class classifier," *Electronic Imaging*, vol. 2018, no. 7, pp. 214–1, 2018.
- [22] E. R. Bartusiak, S. K. Yarlagadda, D. Güera, P. Bestagini, S. Tubaro, F. M. Zhu, and E. J. Delp, "Splicing detection and localization in satellite imagery using conditional gans," in *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 2019, pp. 91–96.
- [23] L. Ali, T. Kasetkasem, F. G. Khan, T. Chanwimaluang, and H. Nakahara, "Identification of inpainted satellite images using evalutionary artificial neural network (eann) and k-nearest neighbor(knn) algorithm," in *2017 8th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES)*, 2017, pp. 1–6.
- [24] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling," *IEEE Transactions on signal processing*, vol. 53, no. 2, pp. 758–767, 2005.
- [25] B. Mahdian and S. Saic, "Blind authentication using periodic properties of interpolation," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 3, pp. 529–538, 2008.
- [26] X. Feng, I. J. Cox, and G. Doerr, "Normalized energy density-based forensic detection of resampled images," *IEEE Transactions on Multimedia*, vol. 14, no. 3, pp. 536–545, 2012.
- [27] P. Boccardo, E. B. Mondino, F. G. Tonolo, and A. Lingua, "Orthorectification of high resolution satellite images," *Politecnico di Torino, Dipartimento di Georisorse e Territorio, Torino–Italy*, 2004.
- [28] A. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 758–767, 2005.
- [29] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli *et al.*, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [30] GeoEye, "Usgs eros archive - commercial satellites - orbview 3," <https://doi.org/10.5066/F7J38R0R>, (accessed Oct 24, 2019).
- [31] USGS, "Usgs earthexplorer," <https://earthexplorer.usgs.gov/>, (accessed Oct 24, 2019).
- [32] E. R. Observation and S. E. Center, "Usgs eros archive - digital elevation - shuttle radar topography mission (srtm) 1 arc-second global," <https://doi.org/10.5066/F7PR7TFT>, (accessed Oct 24, 2019).