Automatically drawing vegetation maps using digital time-lapse cameras in alpine ecosystems

A Preprint

Ryotaro Okamoto *
Doctoral Program in Biology
University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577 Japan.
okamoto.ryotaro.su@alumni.tsukuba.ac.jp

Hiroyuki Oguma
Biodiversity Division
National Institute for Environmental Studies
16-2 Onogawa, Tsukuba, Ibaraki 305-8506 JAPAN
oguma@nies.go.jp

Reiko Ide
Earth System Division
National Institute for Environmental Studies
16-2 Onogawa, Tsukuba, Ibaraki 305-8506 JAPAN
ide.reiko@nies.go.jp

May 11, 2022

Abstract

Enter the text of your abstract here.

Keywords alpine ecology · deep learning · ecosystem monitoring

1 Introduction

The effects of climate change on terrestrial ecosystems are particularly significant in alpine regions (IPCC 2007). Alpine vegetation depends on severe climatic conditions such as low temperatures and long snow-covered periods. Thus alpine areas have rare and unique species adapted to the extreme environments. Several studies have reported that recent global climatic changes, e.g., increasing temperatures and reducing snow-covered periods, have accelerated the invasion of non-native species into alpine areas (see Alexander et al., 2016). In Japan, dwarf bamboo (*Sasa kurilensis*) has invaded alpine snow meadows, probably driven by the extension of the snow-free period (Kudo et al., 2011). Also, climate change has affected the growth and phenologies of native species. For example, the growth of dwarf pine (*Pinus pumila*), a dominant species in Japanese alpine regions, has been affected by climatic conditions such as temperature and snowmelt (Amagai et al., 2015). Monitoring and predicting such changes are essential for effective conservation planning. Since the impact of climate change on alpine vegetation varies depending on species and the microhabitats (Kudo et al., 2010), spatially high-resolution monitoring with a wide range is required.

Previous studies have mainly depended on field observations, yet it is hard to cover broad areas in alpine regions due to poor accessibility and severe weather. Satellite, airborne, and Unmanned Aerial Vehicle (UAV) remote sensing methods seem to be alternatives. However, satellite imageries of alpine areas are rarely available due to cloud cover, and the spatial resolution is not enough to observe vegetation changes at the plant community scale. Airborne imageries can obtain high-resolution data, but its cost becomes a bottleneck for frequent monitoring. Although UAV methods have become popular as a cost-effective tool for ecological monitoring (citation), operating UAVs in alpine regions is challenging due to the strong wind and harsh topology.

*https://github.com/0kam

On the other hand, researchers have also utilized automated digital time-lapse cameras mounted on the ground for monitoring green-leaf phenologies in forests (Richardson et al., 2018), grasslands (Browning et al., 2017) and alpine meadows (Ide and Oguma, 2013). Unlike satellite imageries, such cameras provide images free of clouds and atmospheric effects. Also, they can obtain high-resolution (i.e., sub-meter scale) and frequent (i.e., daily or hourly) images at a meager cost. These studies set some regions of interest (ROI) in the images and calculate the phenology index (e.g., excess greenness, Woebbecke et al., 1995) for each ROI. While we can visually interpret the vegetation distribution and its changes from a set of such time-lapse images, few studies have tackled this. This lack seems to be because applying such time-lapse imageries in monitoring vegetation distribution has two technical challenges.

First, unlike multispectral sensors satellites equip, digital cameras can only obtain three bands (Red, Green, and Blue), making it harder to classify the vegetation. Second, since digital time-lapse cameras are mounted on the ground, applying geo-spatial analysis to these images is challenging. In other words, even if we find a vegetation change, we cannot identify its geographic location and analyze it with topological data. Treating ground-based time-lapse images as geographic data is essential for utilizing these in conservation planning.

This study proposes an automated method for drawing vegetation maps and locating vegetation changes with a digital time-lapse camera by solving these two challenges. We aim to use cheap but powerful digital time-lapse cameras in alpine ecosystem conservation.

2 Materials and methods

2.1 Digital time-lapse camera imagery

We used repeat photography data owned by National Institute for Environmental Studies, Japan (NIES). All the images are publically available on NIES' webpage (`https://db.cger.nies.go.jp/gem/ja/mountain/station.html?id=2`). In 2010, NIES installed the digital time-lapse camera (EOS 5D MK2, Canon Inc., 21 M pixels) on a mountain lodge Murodo-sanso (about 2350 m a.s.l., above the forest limit), located at the foot of Mt. Tateyama (3015 m a.s.l.), in the Nothern Japanese Alps. The camera takes one photograph per hour, from 6 a.m. to 7 p.m. . The camera's field of view (FOV) includes Mt. Tateyama, which ranges from about 2350 m a.s.l. to 3015 m a.s.l. in elevation. The area has a complex mosaic-like vegetation structure because of its topography, including rocks, cliffs, curls, and moraines. From April to November, the camera has observed the snowmelt and seasonal phenology of evergreen, deciduous dwarf trees (e.g., Pinus pumila, Sorbus sp), dwarf bamboos (e.g., Sasa kurilensis), and alpine herbaceous plants (e.g., Geum pentapetalum, Nephrophyllidium crista-galli). We pulled the images from late summer to late fall of 2010 and 2020.

2.2 Preprocessing

2.2.1 Selecting images

First, we selected images that are suitable for vegetation classification. We choose four days with good weather from late summer to late fall each year (9/19, 10/1, 10/11, 10/23 in 2010. 9/19, 10/2, 10/11, 10/23 in 2020). We used the images of this season because we can separate vegetation from the patterns of autumn foliage coloration. Then we pull four images from 11 a.m. to 2 p.m. each day to avoid the effect of temporal noise such as shadows. Finally, we got 16 images for each year.

2.2.2 Automatic image-to-image alignment

Since images are slightly misaligned with each other due to wind, we aligned them before processing. We implemented the program with Python3 language and OpenCV4 image processing library. First, we set one image of 2010 as the alignment target. Next, we automatically found matching keypoints between the target and other images using AKAZE local feature extractor and K-nearest neighbor matcher. Then, we searched and applied the homography matrix that minimizes the distance between each pair of the matching points.

2.3 Automatic vegetation classification

Next, we classified the pixel time series of repeat photography imageries into vegetations. Because deciduous plants have great differences in autumn phenology among species, researchers have used that information for vegetation classification with satellite imageries (e.g., Tigges et al., 2013, Son et al., 2014, Heupel et al., 2018). In our target area, we can recognize the vegetation type with the pixel time series of autumn (Fig.

2). However, no research has applied this technique to ground-based repeat photography imageries. We developed a deep-learning method to take advantage of high-resolution and frequent repeat photography data. Fig.3 shows the classification process.

2.3.1 Model Architecture

Since ground-based photographs can provide high spatial resolution (about ~ 0.5m in our dataset), we can use leaf texture as additional information for classification. For example, we can see a glossy surface on dwarf bamboos and a mat surface on dwarf pines. Thus, we used a small patch (9x9 pixels square) rather than a single pixel as an input of the model. Therefore, a model input is a time series of image patches. Our model has two components to deal with this data structure (see Fig. 3). First, we extracted each patch's features (e.g., texture) using Convolutional Neural Network (CNN) layers (Fig. 4a). CNN is a neural network specialized in recognizing spatial structures of data, such as images. The CNN part outputs a time series of extracted features. Second, we used Recurrent Neural Net (RNN) layers (Fig. 4b) to classify the temporal patterns of the features extracted by the CNN part. RNN is a neural network specialized in recognizing temporal or sequential data dynamics, such as text and speech. Amongst many variants of RNN, we used Long Short Time Memory (LSTM). Combining CNN and RNN, our model can classify the input image-patch time series considering the colors and textures of each patch and their temporal patterns. This CNN-RNN architecture has also been used for audio classification. Using a Deep Learning method has another positive side effect; mini-batch training. Since we used high-resolution images (16 x 21M pix), the training dataset became so immense that we could not feed them to the model at once: that consumes the computer's RAM too much. However, using Deep Learning methods enables us to provide them in small portions (called mini-batches) to save RAM consumption.

2.3.2 Implementation and model training

We implemented the classifier with Python3 language and PyTorch deep neural network library. All source codes are publically available via GitHub (`https://github.com/0kam/xxxx`).

2.3.3 Dataset preparation

We set 5 classes: dwarf pine, dwarf bamboo, other vegetation, no vegetation, and sky. Using free image annotation software (Semantic Segmentation Editor), an expert prepared teacher data (Fig. x) for classification.

3 Automatic georectification

Then, we developed a novel method to convert ground-based landscape imagery into GIS-ready geographical data. This process is called georectification. Georectification of ground-based images has been a difficult task, and this causes the underuse of potentially rich information in ground-based imageries. In plain words, georectification means aligning images into Digital Surface Models (DSMs) so that every pixel of an image gets a geographical coordinate. We can consider a camera as a function that transforms 3D geographical coordinates (latitude, longitude, height) into 2D image coordinates (locations of each pixel). So estimating the parameters of this function (such as the camera location, pose, and the field of view), we can align an image to a DSM. Usually, georectification has three steps: 1. Finding Ground Control Points (GCPs) in the image. 2. Estimating camera parameters such as camera poses and field of view using GCPs. 3. Mapping the image onto the DSM using camera parameters.

Recently, researchers have developed some georectification methods to use ground-based photographs in glaciology (Messerli and Grinsted 2015) and snow cover studies (Portnier et al. 2020). Especially, Portnier et al. 2020 is worth mentioning for its semi-automatic method using mountain silhouettes as GCPs. However, this silhouette-based method has a drawback in the projection accuracy. It only uses limited areas (silhouettes) of images in the image-to-DSM alignment, and also it ignores lens distortion. Because our target site has a complex vegetation distribution and our camera has considerable lens distortion, we needed a more accurate method.

3.1 Local-feature-based matching of images and DSMs

To get matching points between images and DSMs on a broader area, we used an airborne image already georectified. Combining an airborne image and a DSM, we rendered simulated landscape images (Fig. x).

Then we got matching points (GCPs) by applying AKAZE local feature matcher to a target image and this simulated image (Fig. x). GCPs have geographical coordinates (from the DSM) and image coordinates (from the image). Our procedure requires the camera's exact location and initial camera parameters to render the simulated image.

3.2 Modeling and estimating lens distortions

We modeled the lens distortion based on Weng et al. 1992 and OpenCV's implementation (Eq. 1). Our model includes radial (k1~k6), tangental (p1, p2), thin prism (s1~s4) distortion, and unequal pixel aspect ratio (a1, a2). See Weng et al. 1992 for lens distortion modeling. Also, we recommend readers to Portnier et al. 2020 for camera models without lens distortions. Thus, now we can project GCPs' geographical coordinates into image coordinates using lens distortion parameters and other camera parameters (the camera location, pose, and the field of view). We optimized these (except camera location) by minimizing the square projection error of the GCPs using the Covariance Matrix Adaptation Evolution Strategy (CMA-ES). We could not estimate the camera location because it makes the problem too complicated (e.g., a telephoto taken from a distance and a wide-angle taken from a close look similar).

$$\begin{bmatrix} x'' \\ y'' \end{bmatrix} = \begin{bmatrix} x' \frac{1+k_1 r^2+k_2 r^4+k_3 r^6}{1+k_4 r^2+k_5 r^4+k_6 r^6} + 2p_1 x' y' + p_2(r^2+2x'^2) + s_1 r^2 + s_2 r^4 \\ y' \frac{1+a_1+k_1 r^2+k_2 r^4+k_3 r^6}{1+a_2+k_4 r^2+k_5 r^4+k_6 r^6} + p_1(r^2+2y'^2) + 2p_2 x' y' + s_3 r^2 + s_4 r^4 \end{bmatrix}$$

3.3 Implementation and data set

We implemented the algorithm with Python3 language and published it as an open-source package via GitHub (`https://github.com/0kam/alproj`). You can try it with your data. We used an airborne photograph taken in the November of 201X with a spatial resolution of 1.0 m. Also, we used the 5m resolution Digital Elevation Model provided by the Geospatial Information Authority of Japan (GSI) as a DSM.

4 Results

4.1 Vegetation classification accuracy

	Macro Average, N = 5	**Weighted Average**, N = 5	**Dwarf Pine**, N = 5	**Dwarf Bamboo**, N = 5	**Other Vegetations**, N = 5	**Non Vegetation**, N = 5	**Sky**, N = 5
Precision	0.977 (0.008)	0.997 (0.001)	0.977 (0.012)	0.951 (0.013)	0.999 (0.000)	0.959 (0.066)	1.000 (0.000)
Recall	0.987 (0.001)	0.997 (0.001)	0.995 (0.003)	0.960 (0.006)	0.996 (0.002)	0.984 (0.004)	1.000 (0.000)
F1 Score	0.982 (0.005)	0.997 (0.001)	0.986 (0.005)	0.955 (0.004)	0.997 (0.001)	0.970 (0.034)	1.000 (0.000)

4.2 Georectification accuracy

4.3 Vegetation maps

4.4 Detection and localization of vegetation changes

5 Discussion

We suggested a fully automated procedure to transform time-lapse imagery into georeferenced vegetation maps at a meager cost. This task is challenging because of 1. the difficulty of classifying vegetations with ordinal digital camera imagery and 2. the difficulty of georectification. We solved these issues by 1. using the temporal information of autumn leaf colors for vegetation classification and 2. developing a novel method for accurate image georectification. Our vegetation classification performance (with an accuracy of 0.87) and georectification methods (with an RMSE of $rmse$ m) are practical.

5.1 The benefit of using time-series imagery for vegetation classification

One of the shortcomings of ordinal digital cameras is that they only have three bands (Red, Blue, and Green). We made up for this lack of information by using the rich temporal information that time-lapse cameras can obtain. Many plant species have characteristic phenologies, such as flowering and autumn foliages. Observing this requires long-term (such as year-round) monitoring with a high frequency, and digital time-lapse cameras are suitable. This study only focused on the two species (dwarf bamboo and stone pine) that are easy to classify, even with a single image. However, like the previous studies (Tigges et al., 2013, Son et al., 2014, Heupel et al., 2018), we expect our method to classify more vegetations (e.g., dwarf deciduous trees and alpine herbaceous plants) using the temporal information of leaf colors. We also used hourly images for each day. Fig. x shows that this additional information made our model robust against transient noises, such as shadows. Since alpine ecosystems have rough topologies, shadows are a considerable problem. Time-lapse cameras are also beneficial for this reason.

5.2 The performance of our georectification method and its limitation.

5.3 Future application and conclusion