Computers and Electronics in Agriculture 82 (2012) 15–22

Contents lists available at SciVerse ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

A categorical, improper probability method for combining NDVI and LiDAR elevation information for potential cotton precision agricultural applications ☆

Jeffrey L. Willers [a,*], Jixiang Wu [b], Charles O'Hara [c], Johnie N. Jenkins [a]

[a] Genetics and Precision Agriculture Research Unit, USDA, ARS, Mississippi State, MS, United States
[b] Plant Science Department, South Dakota State University, Brookings, SD, United States
[c] Geosystems Research Institute, Mississippi State University and CEO, Spatial Information Solutions, Starkville, MS, United States

ARTICLE INFO

Article history:
Received 29 September 2010
Received in revised form 23 November 2011
Accepted 27 November 2011

Keywords:
Algorithms
Data fusion
Image processing
Improper probability distributions
Categorical variables
Crop management

ABSTRACT

An algorithm is presented to fuse the Normalized Difference Vegetation Index (NDVI) with Light Detection and Ranging (LiDAR) elevation data to produce a map potentially useful for site-specific management practices in cotton. A bi-variate Gaussian probability density distribution is modified to predict an improper probability distribution that also incorporates categorical variables associated with quadrant direction from the population means for the NDVI and elevation data layers. Water availability, influenced by slope and relative changes in elevation (as captured by the elevation data layer), affects crop phenology (as captured by the NDVI data layer). Thus, this fusion procedure results in a map potentially describing the joint effects of NDVI and elevation on cotton growth in a spatial and temporal way.

Published by Elsevier B.V.

1. Introduction

1.1. Background information

Topography, according to several dictionaries, is defined as the arrangement of the natural, artificial, and physical features of an area, as well as a detailed description or representation on a map of such features. The recent availability of remote and proximal sensor systems makes it possible to produce a detailed representation of many types of agricultural topography features and to map them with a particular geographical coordinate system. With the capability of general mixed linear analysis of covariance models (Milliken et al., 2010; Pringle et al., 2010) to demonstrate statistical interactions among the effects of site-specific management decisions, as well with topography covariates (Burris et al., 2010; Willers et al., 2008b), there is an opportunity to develop improved processing methods for collections of various topographical descriptors. Given this usefulness of proximal and remote sensing data layers as descriptors of topography covariates in statistical analyses, their increasing availability raises the question "Can two such data layers be summarized into a single map to simply convey spatial information to an agricultural decision maker?" Such a question presupposes that if site-specific management practices interact with topography variables, then better decisions for crop management should result when an appropriate map combining two influential topography variables is made available to decision makers. The Normalized Difference Vegetation Index (NDVI) or Light Detection and Ranging (LiDAR) data products, when used separately (Willers et al., 2008b) provide some useful geographical information about a crop. Geographically fusing these two topography variables into a single map should produce an even better or a more realistic picture for the agricultural producer or investigator of the status of conditions in a particular field.

1.2. Problem description

The basic problem with combining georeferenced data from different types of sensor systems is variability in measurement units, reference points, and the spectral, spatial, and temporal resolutions. With two topography variables, such as NDVI derived from a multispectral remote sensing system and elevation information derived from a LiDAR system, developing a practical process to fuse them into a map is particularly challenging. The NDVI is a unit-less, continuous number within the interval $[-1,1]$ while elevation is a continuous number expressed in units such as meters (either as

☆ Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendations or endorsement by the US Department of Agriculture.

* Corresponding author. Address: P.O. Box 5367, Mississippi State, MS 39762, United States. Tel.: +1 662 320 7383.

E-mail address: jeffrey.willers@ars.usda.gov (J.L. Willers).

0168-1699/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.compag.2011.11.010

height above the ellipsoid (HAE) or mean sea level (MSL)); therefore, there are also large differences in the data range of each layer.

1.3. Possible solutions

At least two simplistic software approaches are possible solutions. One is to drape the image layer over the digital elevation model (DEM). The second is to load the two image layers into the same viewing window and alter the opacity of the topmost layer. These methods are useful, but the decision maker must interpret the results and extrapolate relationships between the layers. Other approaches involve more effort. For example, one may use the Resolution Merge function of ERDAS® *Imagine* or stack the two input layers with the Spatial Modeler of *Imagine* and then perform (Lillesand et al., 2008) either an unsupervised (ISODATA) or Principal Components classification on the stacked image. In each of these latter methods, a single layer is produced which can be color ramped to display the classification map. The selection of the ramping is arbitrary and the selected ramping options will likely influence the interpretation and thus the usefulness of the output product. Also, the ecological interpretation of these data products is not straightforward which limits their utility for additional statistical analyses.

Other approaches are described in the literature and have provided some insight toward the development of this paper. Edenius et al. (2003) explored how to combine satellite imagery, digital elevation data, and field data to model the distribution of vegetation cover types important to reindeer (*Rangifer tarandus*) during the summer in Scandinavia. These efforts were necessary because of the large land area and a poorly developed infrastructure. In cotton, equivalent constraints exist. Brenning (2009) described another application with respect to the monitoring of rock glaciers, a feature where coarse, blocky debris is cemented by ice a few feet below the surface (Wahrhaftig and Cox, 1959). Brenning sought to improve mapping accuracy by combining process-related terrain attributes derived from digital elevation models and multispectral Landsat TM/ETM+ imagery. He concluded that integrating terrain attributes derived from the DEM and the multispectral image data is necessary for modeling and mapping rock glacier activity. Elaksher (2008) and Koetz et al. (2008) described applications where the goal of the fusion was to improve the classification accuracy of the imagery. Many other applications in the literature appear to use a fusion-type process of digital elevation and digital imagery for the purpose of sharpening the quality or improving the classification of an image layer.

1.4. Objective and outline

This paper has a different goal – the development of a methodology to combine two raster images of agricultural topography that have different informational content to create a fused data product that geographically depicts different regions of cotton growth and vigor. Thus, the objective is to describe an algorithm that fuses two images, a NDVI raster image and elevation raster image of a cotton field, into a new raster image comprised of only a single layer. Following Hogben's (1968) lead, the task includes the succinct use of colors to symbolize these different regions of cotton growth and phenology in the new raster to create a data product that is ecologically interpretable and conducive to statistical analysis.

The organization of the paper is as follows: In 'Section 2', an overview of the two raster products used as inputs to the algorithm is presented. Next is a very brief introduction of two types of coordinate systems used in the development and application of the algorithm. Then, using an early paper by Strahler (1980) as the starting point, the paper provides a short, general description of the development of the algorithm. In 'Sections 3 and 4', a validation of the algorithm is presented by making a comparison to an ISODATA classification (Lillesand et al., 2008) of the same agricultural landscape, but which uses only the multispectral bands. Some potential applications of the resultant data product in cotton management are proposed. Finally, the paper closes by indicating some possible directions for further investigation.

2. Materials and methods

2.1. Image acquisitions and processing

The multispectral data products used in this research were provided through the courtesy of Perthshire Farms, Gunnison, MS, and InTime, Inc., Cleveland, MS (http://www.gointime.com/). The red and near-infrared (nir) bands were employed to determine the NDVI with a ground spatial distance (GSD) of 2.0 m (e.g., Fig. 1). A. Zusmanis, (ERDAS®, Inc., personal communication), provided information on the advantage of taking the arc tangent (ATAN) of the NDVI values to enhance the contrast between bare soil pixels and vegetation pixels. It was calculated and proved useful. The initial LiDAR data product was derived as described in Willers et al. (2008a). This digital surface model (DSM) (Fig. 2) describes elevational relief of cotton fields in meters (m) above the ellipsoidal height (HAE). The GSD of the DSM was 0.5 m.

2.2. The two Cartesian coordinate systems employed by the algorithm

Two different coordinate systems were employed in the data processing. While both are Cartesian coordinate systems, one is an algebraic system (Pignani and Haggard, 1970) involving only the attributes of the two input raster images, which represents the data space (Berry, 1998; Hargrove and Hoffman, 1999) of the rasterized features, elevation and NDVI. The second coordinate system is a mapping system (Langley, 1998) describing the geographic location of each image pixel, in the geographic space (Berry, 1998;

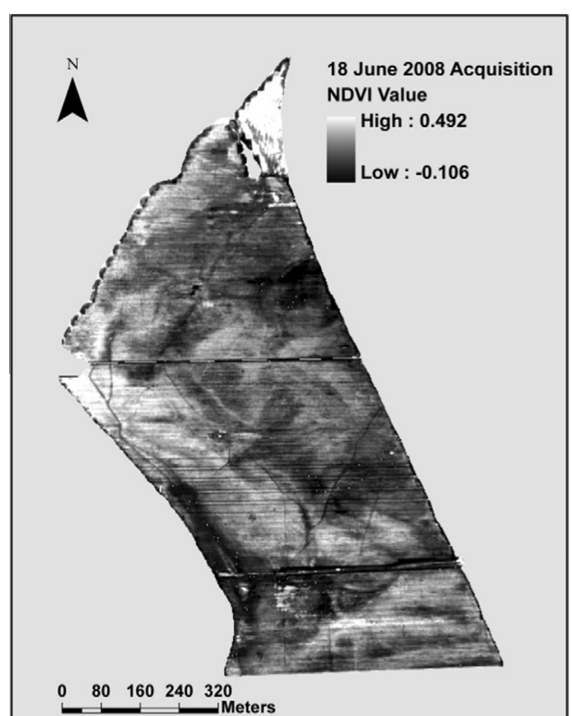

Fig. 1. Normalized Difference Vegetation Index (NDVI) layer of a cotton field landscape for 18 June 2008.