Crop Type Classification by Simultaneous Use of Satellite Images of Different Resolutions

Mark W. Liu, Mutlu Ozdogan, and Xiaojin Zhu

***Abstract*—Accurate and timely identification of crop types has significant economic, agricultural, policy, and environmental applications. The existing remote sensing methods to identify crop types rely on remotely sensed images of high temporal frequency in order to utilize phenological changes in crop reflectance characteristics. However, these image sets generally have relatively low spatial resolution. This tradeoff makes it difficult to classify remotely sensed images in fragmented landscapes where field sizes are smaller than the resolution of imaging sensor. Here, we develop a method for combining high spatial resolution (high-resolution) data with images with low spatial resolution but with high time frequency to achieve a superior classification of crop types. The solution is implemented and tested on both synthetic and real data sets as a proof of concept. We show that, by incorporating high-temporal-frequency but low spatial resolution data into the classification process, up to 20% of improvement in classification accuracy can be achieved even if very few high-resolution images are available for a location. This boost in accuracy is roughly equivalent to including an additional high-resolution image to the temporal stack during the classification process. The limitations of the current algorithm include computational performance and the need for ideal crop curves. Nevertheless, the resulting boost in accuracy can help researchers create superior crop type classification maps, thereby creating the opportunity to make more informed decisions.**

***Index Terms*—Agriculture, classification algorithms, cost function, crops, geospatial analysis, image processing, pattern recognition, probability, remote sensing, vegetation mapping.**

I. Introduction

ACCURATE AND timely information on the location and distribution of major crop types has significant economic, agricultural, policy, and environmental implications [1]. Remote sensing, either by itself or in combination with ground surveys, has been used in crop acreage assessment for over three decades [2]–[5]. While all dimensions of remotely sensed data are relevant, it is the temporal dimension that has been most useful for identifying major crop types [3], [6]–[11]. This is because, at any point during the growing season, crops are at different stages of maturity, manifested as differential spectral response in remotely sensed images to build a crop-specific temporal record. Despite the long history and the promise of temporal monitoring of crop types, remote sensing has not been widely operational for crop acreage assessment. One reason is that, in many places, agricultural fields are relatively small (less than 1 ha) and require image data with medium to high spatial resolution (high-resolution) (5–100 m/pixel) pixels to resolve individual fields [12]. However, this spatial detail comes at the cost of reduced temporal availability. Due to predetermined acquisition strategies and obstructions by clouds, only a few high-resolution images are usually available during critical growing periods. As an alternative, images with lower spatial resolution (low-resolution) (greater than 100 m/pixel) pixels provide frequent coverage but lack the ability to reveal details about the fields. Thus, many of the pixels generated by coarse-resolution sensors are not characteristic of any one crop but comprise a mixture. The purpose of the research presented here is to overcome this resolution tradeoff issue by developing a classification algorithm that merges information from high-resolution low-time-frequency (a few during the growing season) observations with low spatial resolution and high-temporal-frequency (minimum of once per week) observations to map agricultural crop types at high-resolutions.

Many techniques for using multitemporal information for crop type mapping have been developed in the last 40 years. In an early attempt, Smith and Ramey [6] estimated cropped areas using a temporal profile of a vegetative index. Later, Badhwar [3] estimated crop proportions based on at least two images taken at different times of the year. Furthermore, Hall and Badhwar [7] were able to show how temporal profiles of crops could be extended across many geographic regions to assist with crop monitoring. Similarly, Price *et al.* [8] mapped land cover types using multidate Landsat Thematic Mapper (TM) modeling. In India, crops were successfully classified using another multitemporal approach [13]. Spectral matching methods in the spectral and temporal dimensions were also used to map crop types and their irrigation status [5]. The U.S. Department of Agricultural (USDA) National Agricultural Statistics Service (NASS) operationally produces crop data layers at fairly high-resolutions using Landsat data [14], [15].

Generating these high-resolution maps of crop types requires access to several high-resolution images over the growing season. However, given the presence of clouds and acquisition strategies of satellite observations, typically, very few of these images are available during critical periods. When the temporal availability of high-resolution imagery is limited, particularly in areas with small field sizes (less than 1 ha), the concept of

Manuscript received September 7, 2012; revised March 21, 2013 and June 21, 2013; accepted July 12, 2013. Date of publication September 5, 2013; date of current version March 3, 2014. This work was supported in part by the National Aeronautics and Space Administration Application Sciences Program under Grant NNX09AH94G subawarded to Dr. Mutlu Ozdogan.

M. W. Liu was with the University of Wisconsin–Madison, Madison, WI 53726 USA. He is now with 8 to 18 Media, Chicago, IL 60614 USA (e-mail: markwayneliu@gmail.com).

M. Ozdogan is with the Department of Forest and Wildlife Ecology and the Nelson Institute for Environmental Studies, University of Wisconsin–Madison, Madison, WI 53726 USA (e-mail: ozdogan@wisc.edu).

X. Zhu is with the Department of Computer Sciences, University of Wisconsin–Madison, Madison, WI 53726 USA (e-mail: jerryzhu@cs.wisc.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2013.2274431

temporal unmixing of high-frequency observations has been proposed [16]. Instead of classifying each coarse-resolution pixel as a single crop, an unmixing technique is used to recover the fractional area of each crop in a pixel based on its contribution to the mixed response observed by the sensor. For example, Quarmby *et al.* [17] developed a linear mixture approach applied to coarse-resolution sensor data to extract fractional coverage of crop types. More recently, Lobell and Asner [18] showed that a probabilistic approach to temporal unmixing might provide improved results. Moderate Resolution Imaging Spectroradiometer (MODIS) time series data could be used to classify crops in the central plains of the U.S. [10]. Using the temporal information only, Thenkabail *et al.* [19] mapped the irrigation status of crops at global scales. The independent component analysis method was applied to MODIS time series data to automatically map crop types [11]. These methods have been useful for determining the statistical distribution of crops within a given area but are not ideal at generating high-resolution maps of crop types that show the spatial distributions of individual fields that biophysical or economic analyses may require.

Even though the utility of remote-sensing-based crop type classification methods has long been recognized, the methods have rarely transitioned into operational monitoring for a number of reasons. First, most crop type maps derived from remote sensing are only available at coarse resolutions, often containing mixtures of different crops and noncrop surfaces [11], [20]. Second, locations with smaller fields or complex field arrangements require high-resolution observations, but temporal methods retaining the spatial detail are rare. Third, the accuracy of the crop type maps may not be at the desired levels for most operational uses—experience suggests that the accuracy of crop type maps derived from remotely sensed inputs does not exceed 80% [15]. Finally, even with accurate crop type maps, there is still the possibility of errors in area estimation that arise from class definition thresholds [12]. This problem is most immediately noticeable at coarse resolutions but could also occur at fairly high resolutions in locations with smaller (less than 0.5 ha) field sizes [12].

To improve classification accuracy in areas where field-level information on crop types is required, we propose a technique that simultaneously incorporates both high- and low-resolution images. Note that, while there is growing body of literature on combining high- and low-resolution data for land cover mapping [21], to our knowledge, the work presented here is the first attempt at using data from multiple sensors simultaneously in agricultural applications.

II. Methodology

A. Problem Formulation

Existing high-resolution image data are powerful enough to resolve individual fields but are either acquired during that part of the growing season when the crops of interest are least distinguishable or acquired only once. Then, the research question is as follows: Can we develop a methodology to combine information from a high-resolution sensor acquired only a few times during the growing season with low-resolution image data that are frequently available in order to better distinguish the crops of interest?

Ideal Crop Curves: The first step is to generate what we call *ideal crop curves*. These curves contain idealized spectral reflectance values of a crop that vary over the course of the year as the crop goes through its growth cycle. Here, we assume that ideal crop reflectance curves of every crop of interest are available for every spectral band for the sensors being tested. Note that they need not always be the product of observation but can also be modeled using curve fitting [22] or with mechanistic crop growth models [23]. We also assume that the spectral characteristics, including the location and the number of bands in both the high- and low-resolution images, are similar. Using a simple 1-D interpolation function, the observed reflectance values from across the year can be converted into a simple function expressed as $C_Y(t)$, where Y is the crop type and t is the time of year. The output of the function is the expected reflectance value for a crop at a given time. For now, we simplify the problem by assuming that there is only one spectral band of interest. As will be shown later, incorporating the other spectral bands into the problem formulation is quite straightforward.

Given a set of images containing high- and low-resolution pixel data, the input data may be referred to as follows:

$X_{ij}^{\mathrm{hi}}(t)$ observed reflectance value for the high-resolution image at pixel ij at time t;
$X_{ij}^{\mathrm{low}}(t')$ observed reflectance value for the low-resolution image at pixel ij at time t'.

Note that we use (t) to denote time series of high-resolution images and (t') to denote time series of low-resolution images, which may or may not overlap in time with the high-resolution observations. We also use function $C_Y(t)$ to predict a reflectance value as a function of time for a particular crop Y. For a simplified version of the problem that involves only a single spectral band, the variables are further defined as follows:

Y_{ij} predicted crop type at pixel ij;
$C_{Yij}(t)$ ground reference reflectance value for pixel ij at time t assuming crop type Y.

Now, the problem can be reduced to an objective function that must be minimized. The goal is to solve for the best possible Y_{ij} for all pixel values ij of interest. The function $C_{Yij}(t)$ is entirely dependent on the prediction for the matrix Y and the ideal curves that are being used.

Optimizing Using Only the High-Resolution Images: The first function that must be minimized is the squared difference between each predicted pixel and observed pixel for the high-resolution images using a typical least squares method expressed by

$$\sum_{(i,j),t} \left[C_{Y_{ij}}(t) - X_{ij}^{\mathrm{hi}}(t)\right]^2 \tag{1}$$

which states that the goal is to find the crop type Y_{ij} for each pixel that has a crop curve that best fits the observed data for that pixel over time.

Optimizing Using Low-Resolution Images: Next, the differences in magnitudes between predicted and observed pixel

values for the low-resolution images must be minimized. However, the scale of the predicted grid of crop types and the scale of the low-resolution observations will not match as the predicted grid will have many more pixels in the matrix. To compare these values, a function can be created to enable several smaller pixels to combine to create a single pixel with coarser resolution. However, most sensors do not simply average values over a large grid but contain weights that vary across the pixel, defined by the sensor's point spread function (PSF). A PSF describes a probability density function around the center of a pixel [24] and assuming that all low-resolution pixels have identical PSFs. Using a Gaussian distribution to model a sensor PSF

$$\omega\left((i,j)(k,l),\sigma_p^2\right)=\frac{1}{\sigma_p\sqrt{2\pi}}e^{-\left((k-i)^2+(l-j)^2\right)/\left(2\sigma_p^2\right)} \quad (2)$$

where (i,j) is the center of the high-resolution pixel, (k,l) is the center of the low-resolution pixel, and σ_p is the standard deviation of the Gaussian PSF. Note that both (i,j) and (k,l) are defined in the same coordinate system. The Gaussian PSF simply states that the high-resolution pixels that are nearer to the center of the corresponding low-resolution pixel would carry a larger weight in determining the value for the low-resolution pixel. Thus, the weight $\omega_{(i,j),(k,l)}$ is just the probability density function evaluated on the points (i,j) and (k,l) using a predefined variance σ_p^2.

With the definition of PSF in place and assuming a least squares method, the role of the second function to be minimized is to describe how close the reflectance value of the low-resolution image pixel is to the weighted sum of the reflectance values for the corresponding high-resolution pixels

$$\sum_{(k,l),t'}\left[\frac{\sum\limits_{(i,j)\in\Omega_{kl}}\omega_{(i,j),(k,l)}\cdot C_{Y_{ij}}(t')}{\sum\limits_{(i,j)\in\Omega}\omega_{(i,j),(k,l)}}-X_{kl}^{\text{low}}(t')\right]^2. \quad (3)$$

The denominator in the first term is used to normalize the weights in order to make them sum up to one. A summary of the variables used in (3) is as follows:

$X_{kl}^{\text{low}}(t')$	observed reflectance value for the low-resolution image at pixel kl at time t;
σ_p^2	variance of the Gaussian PSF;
Ω_{kl}	area within two standard deviations of pixel center (k,l);
$\omega_{(i,j),(k,l)}$	weight of high-resolution pixel (i,j) with respect to low-resolution pixel (k,l).

See the data section hereinafter for setting σ_p^2 for synthetic and real data. Again, t, the high-resolution time stamp, is less frequent than the low-resolution time (t') and may not coincide with it in time.

Optimizing Using Spatial Regularization: The third function to consider is a spatial regularizer based on the concept that adjacent pixels are more likely to have the same crop type simply because of geographic proximity. This is reasonable because two adjacent high-resolution pixels will often belong to the same crop type in the same field. The regularizer will simply count the number of adjacent pixel pairs that have different predicted crop types. A higher number of pairs of adjacent pixels that are different will lead to a higher penalty

$$\sum_{(i,j)\text{ with neighbor }(p,q)}1(Y_{ij}\neq Y_{pq}). \quad (4)$$

The $1()$ indicator function in (4) evaluates 1 if $Y_{ij}\neq Y_{pq}$ and evaluates 0 if $Y_{ij}=Y_{pq}$.

Combining the Optimization Functions: We now minimize each of the three equations simultaneously by choosing the optimal matrix of predicted crop types Y

$$\min_{\{Y\}}\left[(1)+\lambda_1(3)+\lambda_2(4)\right] \quad (5)$$

where the λ_i's correspond to the nonnegative weights, determined *a priori* by the user, for each function being minimized.

Another way to express this minimization function is to define a maximum likelihood function that expresses the likelihood of every possible matrix Y: The higher the likelihood, the better the predicted crop type matrix Y describes the data. Strictly speaking, this can be classified as a *maximum a posteriori* problem where the first two terms $[(1)$ and $\lambda_1(3)]$ represent the log likelihood and the third term $[\lambda_1(3)]$ is not a likelihood but rather a log prior

$$P(Y|X^{\text{hi}},X^{\text{low}})\equiv\frac{1}{z}e^{-((1)+\lambda_1(3)+\lambda_2(4))}. \quad (6)$$

Here, z is simply a normalizing constant to ensure all of the likelihoods sum to unity over all possible predicted crop matrices Y

$$z=\sum_{\{Y\}}e^{-((1)+\lambda_1(3)+\lambda_2(4))}.$$

The optimization problem in (5) minimizes error over discrete variables Y, known to be difficult as an example of combinatorial optimization. By relaxing the problem into a continuous optimization problem, a wide range of well-established optimization algorithms to find a near-optimal solution can be utilized. In fact, these optimization algorithms in continuous space are crucial in developing a practical implementation of the crop type classifier.

In continuous space, each pixel will now consist of a vector of probabilities that describe the strength of belonging to each crop type. For instance, for three crops such as maize, wheat, and barley, each pixel in the discrete space would be labeled as one of them. By contrast, in continuous space, each pixel would be labeled as having a probability of being each of these types. A particular pixel might be labeled as (0.7, 0.2, 0.1), meaning 70% likely to be maize, 20% wheat, and 10% barley. Once these probabilities are solved, the most likely crop type (the entry with the largest probability) is chosen to label that pixel.

To relax the optimization problem from discrete space to continuous space, the matrix of predicted crop types that are being solved for, Y_{ij}, must be redefined. More specifically, now, Y_{ij} must be redefined to be a vector that expresses the probability of pixel (i,j) being each crop type. Thus, if there

are n possible crop types, each element in the matrix Y can be defined as follows:

$$Y_{ij} = \begin{pmatrix} Y_{ij}1 \\ Y_{ij}2 \\ \cdots \\ Y_{ij}n \end{pmatrix}.$$

Note that the entries are nonnegative and sum to unity. In addition, each entry in $CY(t)$ must now return a vector of predicted reflectance values for pixel. Each element in these rows corresponds to the predicted reflectance value for each crop at a time t. To this end, we define

$$C(t) = (C_1(t)C_2(t)\cdots C_n(t))$$

and replace each instance of $C_{Yij}(t)$ with $C(t)' \cdot Y_{ij}$, where the dot here is element-wise in the product.

In the continuous case, the optimization algorithms take the following form. For optimizing using high-resolution pixels

$$\sum_{(i,j),t} \left[C(t)' \cdot Y_{ij} - X_{ij}^{\text{hi}}(t)\right]^2 \tag{7}$$

and for optimizing the low-resolution pixels

$$\sum_{(k,l),t'} \left[\frac{\sum_{(i,j)\in\Omega_{kl}} \omega_{(i,j),(k,l)} \cdot C'(t') \cdot Y_{ij}}{\sum_{(i,j)\in\Omega_{kl}} \omega_{(i,j),(k,l)}} - X_{kl}^{\text{low}}(t') \right]^2. \tag{8}$$

The spatial regularizer must also change from discrete to continuous space. Instead of penalizing each pair of pixels with an indicator function, the new penalty imposed will be the vector norm between the label vectors of the two pixels. Thus, if the vectors are similar to one another, the penalty will be small, but if the pixels are very different, the penalty will increase

$$\sum_{(i,j)\text{ with neighbor }(p,q)} \|Y_{ij} - Y_{pq}\|. \tag{9}$$

Finally, combining these three optimization functions, the objective function that must be minimized is

$$\min_{\{Y\}} [(7) + \lambda_1(8) + \lambda_2(9)] \tag{10}$$

subject to :

$$Y_{ijk} \geq 0 \qquad \forall ij \quad \forall k = 1, \ldots, n$$
$$\sum_{k=1}^{n} Y_{ijk} = 1 \qquad \forall ij.$$

It is important to note that, in (10), the objective function is convex, making it is easier to optimize.

The optimization approach that we selected was the *projected gradient descent* (PGD) algorithm [25]. PGD is an iterative approach where each successive step brings the algorithm closer to a solution, relying on the gradient of the objective function being minimized. It uses this gradient to determine the optimal direction to move the solution with each successive iteration. To determine a sufficient number of iterations, two parameters must be set: The parameter *stepSize* determines how large a step that the algorithm should take in the appropriate direction, and *minChange* indicates when the algorithm can be considered to have converged. Finally, because the optimization problem is constrained, a projection step is performed if the gradient descent step moves the solution outside the feasible region (i.e., violating any of the constraints). This is done by projecting the solution back to the closest point in the feasible region. To project y^i back to the feasible region, we used the simple heuristic of setting all negative elements in the vector to zero and then normalizing all of the elements so that they sum to one. When this algorithm finishes, y^i is the solution.

B. Implementation

Working With Synthetic Data: For testing with synthetic data, we generated a 2-D array of finite dimensions where each element in the array is a labeled crop type. The dimensions of this array are identical to the dimensions of the number of pixels in the desired high-resolution image. In the crop array, we eliminated bias toward any particular crop type by ensuring that the probability of each crop type being represented in the map is the same as the truth map, even though this is unlikely to be true in real scenes. Next, the size of each field was selected to be slightly larger than a single high-resolution pixel. Fields that are between 1 and 20 pixels are all reasonable sizes for an individual field of a particular crop. To more accurately represent real scenarios, the fields should differ in size and should be arranged in an asymmetrical pattern. Thus, field sizes should be fairly small, and the shapes and relative positions of these fields should be somewhat irregular. The resulting synthetic crop landscape was used as a reasonable representation of a real field.

To create the 2-D array of reflectance values that represent the synthetic multispectral image, every individual pixel's crop type label is examined, and a reflectance value is assigned by using the ideal crop curves. We first determine what the reflectance value would be at that time of the year that the high-resolution image represents if it were the perfect crop and then add some noise to that reflectance value. For the purposes of this exercise, we extracted these ideal crop curves using MODIS data by first finding homogeneous land cover units in higher resolution images and reading temporal changes in spectral reflectance values in each band from visible to shortwave infrared portions of the electromagnetic spectrum.

After finding the ideal reflectance value for a particular pixel, zero-mean Gaussian noise is added to further approximate real curves. The standard deviation of the Gaussian distribution was chosen to create synthetic data that were reasonably difficult for standard classification algorithms to classify. Assuming there are between one and three high-resolution images over the course of a year for the synthetic data, a good choice for one standard deviation for the Gaussian distribution is 0.15. This is substantial noise, considering that the range of possible reflectance values lies between zero and one, and the majority of reflectance values are below 0.6. Addition of noise allows us to explore various ways that the ideal crop curves could be different across fields. First, farmers may plant crops at different times of the year, causing the reflectance curves to be shifted in time. Second, there is significant variability in the reflectance

values of crops even if they are planted at the same time of the year. Individual plants or fields may grow at unique rates based on the amount of irrigation, the quality of the soil, the amount of sunlight, the planting density, and any number of other factors. Third, there may be sensor noise while using a real sensor that would also influence an ideal reading.

To create the low-resolution synthetic image, we chose pixel resolution to be equal to the size of 64 high-resolution pixels. Thus, the width and height of each low-resolution pixel are eight times greater than the width and height of each high-resolution pixel. Since the crop type label of each of the 64 high-resolution pixels is known, the low-resolution pixel's multispectral reflectance value over time was calculated by first entering the date of the low-resolution image into the ideal crop curve for each crop type within the low-resolution pixel and weight these reflectance values, based on the distance between the center of the low-resolution pixel and the center of the high-resolution pixel in question. Finally, Gaussian noise with zero mean and the same high-resolution standard deviation (0.15) is added to the estimated low-resolution ideal curves.

Working With Real Data: While the choice of sensors with different spatial/temporal characteristics is determined by the user, the new algorithm should, in theory, work with any pair of images with different spatial resolutions and temporal acquisitions. A more important constraint is the location and number of spectral bands. The algorithm requires that both sensors gather data in similar spectral bands. Moreover, there must be sufficient number of bands in common between the sensors, although the fewest common bands that the sensors need for the algorithm to be effective are not yet known. The temporal resolutions of the sensors are important as well. If the sensors have similar temporal resolutions, then the algorithm is somewhat ineffective as its purpose is to feed high-temporal-frequency observations from a coarse-resolution sensor into a set of fewer observations at higher spatial resolution during the classification process. Hence, the methodology proposed here works best when only a small number of high-resolution images are available during the growing season to be paired with a relatively large number of low-resolution images from the same period.

Landsat and MODIS data complement each other well because Landsat produces images every 16 days while MODIS produces images every 1–2 days. Furthermore, we know that Landsat and MODIS have six spectral bands in common [26].

For real images, we chose an area that was familiar and had an abundance of both usable remotely sensed data and truth maps. The site is in the Mississippi River Basin (centered at 33.67 N and −90.45 W). The high-resolution images were from the Landsat TM and Enhanced TM Plus sensors, while the low-resolution images were from the MODIS Nadir bidirectional reflectance distribution function (BRDF) Adjusted Reflectance (NBAR) product [MOD43] eight-day composited data set [27]. Each Landsat image has a spatial resolution of 30 m and covers a 20 by 20 km region, stored in the Universal Transverse Mercator (UTM) projection [Fig. 1(a)]. All images were from the year 2002. The Landsat data have six bands—Band 1 (blue, 450–515 nm), Band 2 (green, 525–605 nm), Band 3 (red, 630–690 nm), Band 4 (near infrared (NIR), 780–900 nm), Band 5 (1550–1750 nm), and Band 7 (2090–2350 nm). The images were atmospherically corrected using the Landsat Ecosystem Disturbance Adaptive Processing System approach [28]. Table I provides the details of each of these data sets along with the growth stages of crops considered. The fields in this region were larger than the Landsat pixel size but much smaller than the MODIS pixels, which was ideal for the proof of concept.

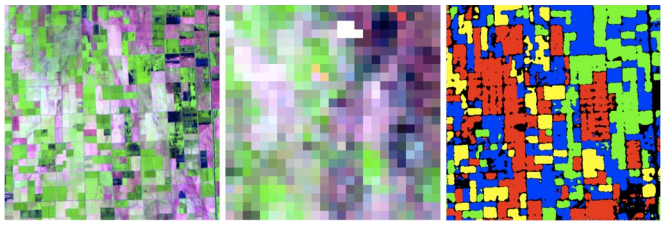

Fig. 1. Satellite images and the reference map of the test site. (a) Bands 7–4–2 as red–green–blue (RGB) color composite of the Landsat image on July 4, 2002. (b) Bands 7–2–4 s RGB color composite of the MODIS image. It is an eight-day composite from June 9 through June 17. We selected this image rather than the one closer to July 4 because this was the raw image closest in time that was free of clouds. (c) Crop type labels for the real data. Yellow = Corn, Red = Cotton, Green = Rice, Blue = Soy, Black = Unidentified/Background. The data were extracted from USDA NASS cropland data layer (NASS, 2011).

TABLE I
INFORMATION ON LANDSAT IMAGES USED IN THIS STUDY. ALSO SHOWN ARE THE CROP STAGES THAT CORRESPOND TO THESE DATES

Image date	Crop stage
January 25	fallow
February 18	fallow
March 22	early rice planting, early corn planting
July 4	cotton mid-season, corn mid-season, rice mid-season, soybean mid-season
August 21	rice early harvest, corn mid-season, cotton mid-season, soybeans mid-season
August 29	rice early harvest, corn mid-season, cotton mid-season, soybeans mid-season
September 6	rice early harvest, corn mid-season, cotton mid-season, soybeans mid-season
September 22	Corn early harvest, cotton early harvest, soybean harvest
October 16	cotton harvest, corn harvest, soybean harvest

The NBAR product comes with 500-m resolution pixels projected on the sinusoidal 10° grid. We reprojected these data to a UTM grid to match the Landsat observations [Fig. 1(b)]. Because the date of imagery is an important input into our algorithm, we chose to use the actual day(s) of compositing when representing the timing of MODIS imagery with respect to Landsat images. This eliminated the confusion of which day(s) to use within the eight-day compositing period [29].

The reference crop type map was generated by the USDA NASS, using a supervised decision tree classification algorithm [15]. The map contains four crop types: corn, cotton, rice, and soybeans [Fig. 1(c)]. The black areas represent the areas where either no crop was identified or noncultivated areas.

To generate the ideal crop curves, we first isolated pixels belonging to each of the crop types in the NASS map by masking. We then aggregated these single crop maps to match the size of MODIS pixels using a cubic convolution resampling. Pixels in the aggregated map that had greater than 90% cover for the crop of interest were labeled as pure crop pixels at MODIS scale. We extracted temporal profiles of surface reflectance data across all

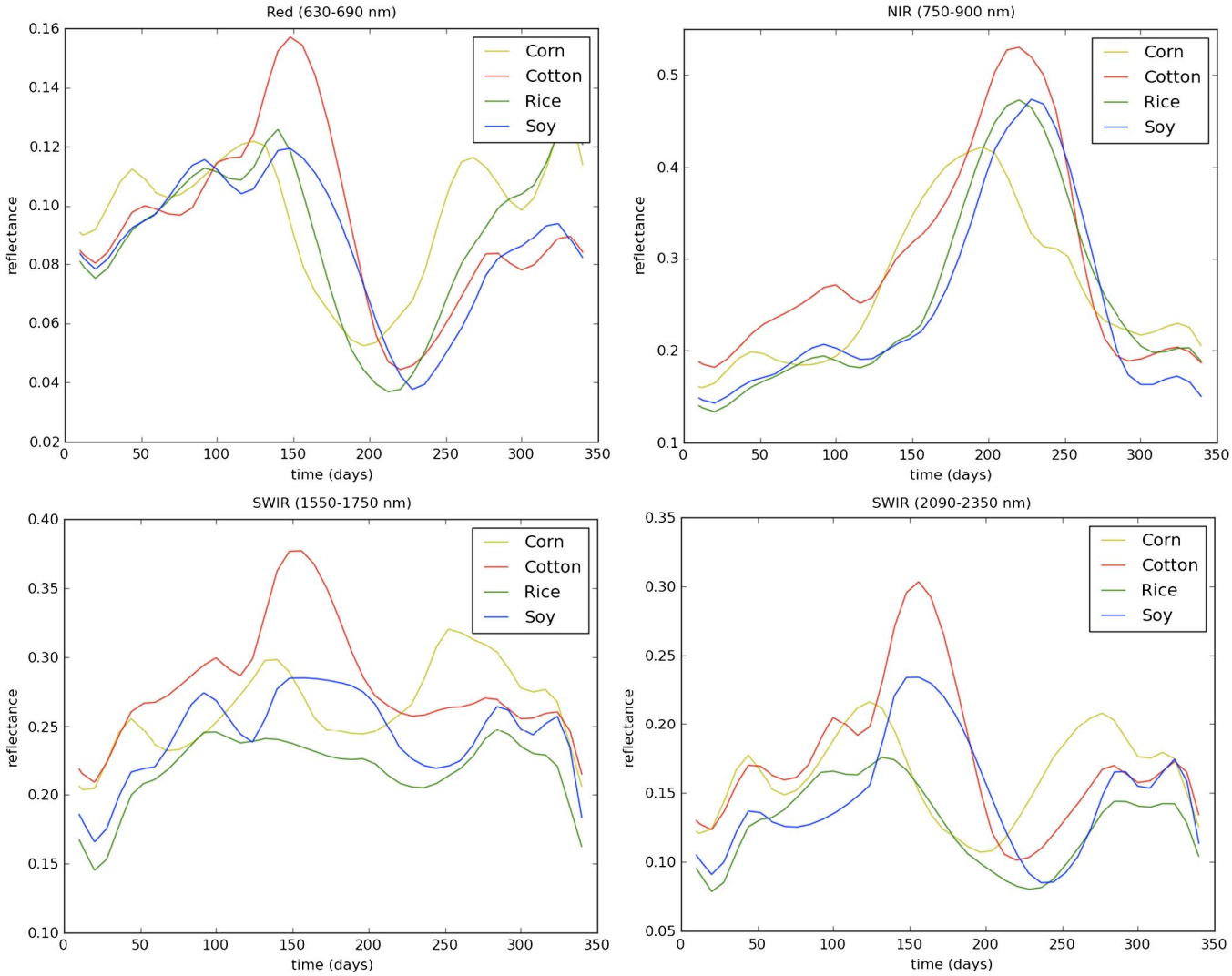

Fig. 2. Temporal profiles (ideal crop curves) of four different crops in (upper left) red, (upper right) NIR, (lower left) SWIR, and (lower right) SWIR portions of the electromagnetic spectrum extracted from the MODIS data by interpolation.

MODIS bands using only these pure crop pixels and used their average as the ideal crop curves. Depending on the crop type, the number of temporal profiles ranged from 10 to 25 for the study area (Fig. 2). Note that, in some cases, the data set used to generate the ideal crop curves covers a different area than the area being classified.

Tuning Algorithm Parameters: To tune the neighbor lambda coefficient described by the objective function, we chose an arbitrary set of training data, and the accuracy of the gradient descent classification was plotted as the value of the neighbor lambda increased. Thus, the training data consisted of a single image rather than multiple high-resolution choices. Since real images and synthetic ones will have slightly different characteristics, it was important to tune these parameters separately for the synthetic and the real data.

For the synthetic images, a training image was created at day 170, and the noise added to it had a standard deviation of 0.15, where the reflectance values ranged from zero to one. This was a significant amount and thus provided an image that would yield low classification accuracies if classified using only the high-resolution data and no neighbor heuristic or low-resolution data.

When tuning lambda parameter for real data, the image chosen was from the 217th day of the year (August 4, 2002). Numerous tests were run on this image using different values for neighbor lambdas that spanned over four orders of magnitude. The peak was not large but could be seen most clearly in tests that spanned the range from 0 to 0.1.

The low-resolution lambda coefficient was also tuned using only a single high-resolution image. To tune this parameter, the neighbor lambda coefficient was set to zero, and the low-resolution lambda was varied. For all tests, the data set used ten low-resolution images spanning the year. For the synthetic images, the high-resolution training image was day 170 again, with the same amount of noise to tune the neighbor lambda coefficient. There was a steep drop once the low-resolution lambda parameter reaches a tipping point, indicating that, if the low-resolution lambda is set too high, then the entire classification would be decided solely by the low-resolution images, and the information from the high-resolution images would no longer assist the classification.

For real data, the optimal parameter was obtained when varying lambda from 0 to 100. The optimal value was about 60, and this low-resolution lambda value was used in all classifications.

TABLE II
DESCRIPTION OF DIFFERENT EXPERIMENTS WITH REAL DATA INCLUDING THE IMAGE DATES

Experiment number	Low res image dates	Experiment description
1	(no low res images)	High res images only
2	(no low resolution images)	High res + neighbor heuristic
3	2/6, 3/18, 4/19, 5/29, 6/30, 8/9, 9/10, 10/20, 11/21, 12/31	High res + neighbor heuristic + 10 low resolution
4	1/21, 2/6, 2/22, 3/18, 4/3, 4/19, 5/5, 5/29, 6/14, 6/30, 7/16, 8/9, 8/25, 9/10, 9/26, 10/20, 11/5, 11/21, 12/7, 12/31	High res + neighbor heuristic + 20 low resolution
5	1/13, 1/29, 1/6, 2/22, 3/2, 3/18, 3/26, 4/11, 4/19, 5/5, 5/13, 5/29, 6/6, 6/22, 6/30, 7/16, 7/24, 8/9, 8/17, 9/2, 9/10, 9/26, 10/4, 10/20, 10/28, 11/13, 11/21, 12/7, 12/15, 12/31	High res + neighbor heuristic + 30 low resolution
6	1/13, 1/21, 1/29, 2/6, 2/14, 2/22, 3/2, 3/18, 3/26, 4/3, 4/11, 4/19, 4/27, 5/5, 5/13, 5/29, 6/6, 6/14, 6/22, 6/30, 7/8, 7/16, 7/24, 8/9, 8/17, 8/25, 9/2, 9/10, 9/18, 9/26, 10/4, 10/20, 10/28, 11/5, 11/13, 11/21, 11/29, 12/7, 12/15, 12/31	High res + neighbor heuristic + 40 low resolution

C. Testing the Algorithm

To determine how well this algorithm performs after low-resolution data are added, we ran six groups of tests. The first test is simply a traditional maximum likelihood classification algorithm with no enhancements. In the second test, the neighbor heuristic is added using the value for the neighbor lambda that was tuned earlier. In the third test, the neighbor heuristic is also used, but ten low-resolution images spaced evenly across the year are also added. The low-resolution lambda in these tests is the value that we calculated earlier. The fourth, fifth, and sixth tests are similar to the third but use 20, 30, and 40 low-resolution images, respectively. Each of these tests is labeled with an experiment number (see Table II).

Intuitively, high-resolution images chosen at times of the year when the crops to be separated have different reflectance values will allow the classifier to generate better results. For each of the experiments listed in Table II, additional tests were conducted to determine how the classification improves with a different set of high-resolution inputs. Nine sets of high-resolution inputs were used for this, and for each of these, the classification accuracy was plotted to track the level of improvement as more low-resolution images are added (Table III).

The physical size of the synthetic images was fairly small: The high-resolution images were 32×32 pixels, and the low-resolution images were 4×4 pixels, respectively. The small size of the synthetic images allowed multiple iterations of the classification algorithm, where each test was run on 30 independently generated sets of high- and low-resolution images and then averaged.

For real data, the computational and memory requirements of the gradient descent algorithm were much more limiting. The gradient descent algorithm calculates and stores the partial derivatives for every variable of interest, considering that every single pixel has six variables corresponding to six spectral bands. Furthermore, keeping track of all the partials for all of the variables requires that the partial derivatives for every variable with respect to every other variable be known. If a high-resolution image is 400 pixels wide and 400 pixels tall, the number of coefficients that need to be calculated would be $400 \times 400 \times 6 \times 400 \times 400 \times 6$. Assuming that each coefficient is stored in a 4 B, 14.8 TB of memory would be needed to hold this array of coefficients, which is unrealistic. To reduce this burden, we simplified the real data by dividing them into smaller pieces. More specifically, a 400×400 pixel image was first divided into 64-image windows, each 50×50 pixel in size and classified separately, and then reassembled. Despite this simplification, each 50×50 pixel image still took nearly 1 h to classify, even without low-resolution data.

TABLE III
HIGH-RESOLUTION IMAGE CHOICE DATES AND THE EXPERIMENT DESCRIPTIONS

Test group	Image dates	Description of high resolution images
1	3/22	one high resolution image chosen from the first half of the year
2	10/16	one high resolution image chosen from the last half of the year
3	7/4	one high resolution image chosen from the middle third of the year
4	1/25, 3/22	two high resolution images chosen from the first half of the year
5	8/29, 10/16	two high resolution images chosen from the second half of the year
6	3/22, 8/21	two high resolution images: one from the first half of the year, one from the second half of the year
7	1/25, 2/18, 3/22	three high resolution images chosen from the first half of the year
8	8/29, 9/22, 10/16	three high resolution images chosen from the second half of the year
9	3/22, 7/4, 9/6	three high resolution images: one from the first third of the year; one from the middle third of the year; one from the last third of the year

Raw outputs from a classification algorithm often contain significant errors when examined on a pixel-by-pixel basis. To overcome this issue, we employed a postclassification segmentation process to translate initial per-pixel results to per-polygon (segments) outcomes. We used a segmentation algorithm developed in [30] with a minimum mapping unit of six Landsat pixels (roughly equal to 0.5 ha). Input features to segmentation included a combination of Landsat red, NIR, and short-wavelength infrared (SWIR) bands from the peak crop growth period. Per-pixel classification results were first eroded and dilated, and converted to polygon-based outputs using the majority rule. The final polygon-based crop maps were devoid of high-frequency noise often present in per-pixel maps.

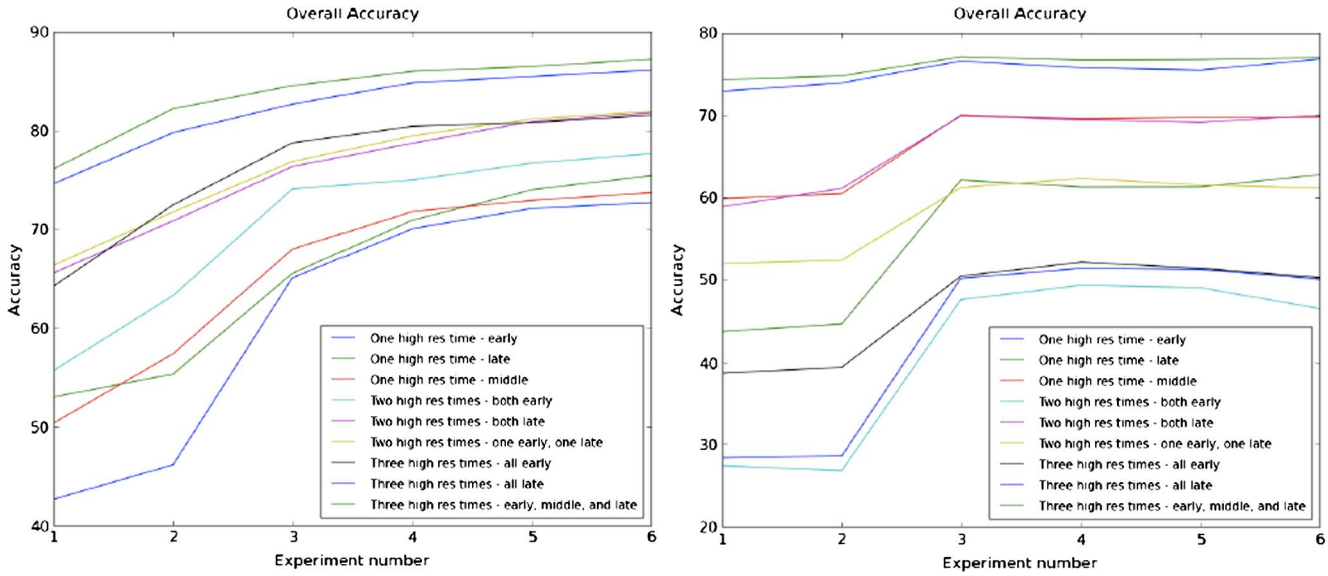

Fig. 3. Overall accuracy of classification results using both (a) synthetic and (b) real images with varying inputs. For the synthetic image results, each line on the graph represents a test group where high-resolution images were chosen at certain times of the year. The lines are labeled with these times. Every data point represents an average of 30 trials with 30 independently generated sets of synthetic data. For real data, as the lines move from left to right, more low-resolution data are added as inputs to the classifier. Each line represents a different set of high-resolution inputs, using the tuning parameters calculated earlier. See Tables II and III for the list of different inputs.

Our goal is not to classify a particular area with the best possible accuracy but rather to see how the classification accuracy for a variety of scenarios can be improved using this new technique. Thus, we assumed that the investigator has access to between one and three high-resolution images and has ten or more low-resolution images during the year. When we ran the tests listed in Table II, the number of low-resolution images was varied.

III. Results

Overall, the proposed algorithm appears to effectively boost crop type classification accuracy when very few high-resolution inputs exist. Whether with synthetic or real data, we found that, if the investigator has only a single high-resolution image, its classification accuracy can be boosted by 10%–20% simply by adding ten low-resolution images during the classification process.

With synthetic data, addition of low-resolution imagery improved crop type classification accuracy by between 10% and 40%, depending on when the high-resolution images were chosen [Fig. 3(a)]. Most gains occurred in cases where the hi-re images were few and acquired early or very late in the growing season. Increasing the number and the distribution across the growing season of high-resolution data alone improved classification by between 10% and 20%; adding the low-resolution data further boosted the accuracy.

With real data, the improvement in classification accuracy is most pronounced when the high-resolution images occur at times of the year when the crop curves most resemble one another [Fig. 3(b)]. For example, when we used either one or two high-resolution images from very early in the year (January 15 and March 22), ten well-spaced low-resolution images boosted accuracy by 22% in the case of a single high-resolution image and by 20% for two images. This is an extremely promising result. It suggests that, if only a single high-resolution image is exists but is accompanied by many low-resolution images, there is the potential to boost classification accuracy by 20% through the use of this algorithm when low-resolution images are incorporated into the classification.

The improvement decreases when the high-resolution images are chosen from times of the year when the crop curves differ substantially [Fig. 3(b)]. When a single image is chosen in mid-year, on July 4, addition of ten low-resolution images to the classifier boosted the accuracy by 10%. This improvement is still desirable but not nearly as beneficial as when the images are from the beginning of the year.

The smallest improvements occurred when the classifier used multiple high-resolution images chosen at those times of the year when the crop curves differed the most. The two tests that showed the least improvement were when there were three high-resolution images spaced across the year (March 2, July 4, and September 6) or when there were three such images all in the second half (August 29, September 22, and October 16). For these two tests, the classification accuracy improvements from the addition of ten low-resolution images to the classifier were 3.1% and 3.4%, respectively.

Note that, with real data, accuracy actually decreases when more than ten low-resolution images are added [Fig. 3(b)]. Eventually, their weight becomes too great; thus, the low-resolution lambda parameter must be decreased appropriately to compensate. This result is analyzed in greater depth in the discussion section.

After completing each classification run with real data, each of the 64 parts was pieced back together into a 400 by 400 pixel classification map that was the same size as the original. The classification maps that use only high-resolution data are placed next to those that use high-resolution data with the neighbor

heuristic and ten low-resolution images. Next to these is the truth classification map for comparison. Similar sets of three images were created for each of the nine groups of high-resolution input images listed in Table III. These images are shown in Fig. 4(a)–(c).

Overall, rice was classified more accurately across all inputs and iterations of the classification algorithm. The primary reason for this is directly related to the temporal/spectral characteristics of rice. Paddy rice spends a part of its development period in water, which is reflected in the temporal reflectance spectra. Also, it has lower SWIR reflectances without a defined peak. Moreover, rice is always classified as being present where it should be, even with many false positives, where other crops are being mistaken for rice.

On the other hand, corn and cotton were very difficult to distinguish. This is partially due to the significant overlap in their spectrotemporal profiles (Fig. 2) and partially due to their position on the landscape. More specifically, corn occupies isolated fields in the truth image while all other crops are planted as groups of fields, perhaps to maximize management and irrigation. When low-resolution data were brought in, areas of isolated fields tended to be less accurate due to smoothing than areas where groups of same crops are located.

To further examine the changes in classification accuracy, we chose a test case where the initial accuracy using only high-resolution images was fairly poor but improved significantly when low-resolution images were added. The classification accuracy of a single high-resolution image from October 16, 2002, with no neighbor heuristic or low-resolution images is summarized in Table IV. Results show that, without any low-resolution images, the accuracy for cotton is fairly poor. Virtually, all cotton fields were labeled corn, with the producer's accuracy near 0.01%. Moreover, while corn's producer's accuracy is very high at 90.29%, its user's accuracy is quite low at 24.53%, suggesting that too many fields were labeled as corn. By examining the ideal crop curve, it can be seen that there are significant differences in the ideal reflectance values for the crops, but a single high-resolution image fails to capture this differently because of its timing. Because the classifier is having such difficulty even though there are such differences in the reflectance values, it is obvious that a single image does not suffice to classify crop types accurately.

When the same high-resolution image was classified with the neighbor heuristic and addition of ten low-resolution images, its overall accuracy improved dramatically, from 43.55% to 62.65% (Table V). Most striking, however, were the changes in the producer's accuracy for cotton, improving from 0.01% to 72.77%. This improvement resulted mainly from the cotton pixels that had been incorrectly labeled as corn now being correctly labeled. Similarly, because fewer of the other classes were now incorrectly labeled as corn, there was a steep increase in the corn's user's accuracy: Without the low-resolution images, its user's accuracy was 24.53%, but after adding the ten low-resolution images, it jumped to 61.58%.

For rice, the user's and producer's accuracies both increased significantly after adding ten low-resolution images. The user's accuracy improved from 68.83% to 76.97%, and the producer's accuracy increased even more, from 52.95% to 70.11%. Many

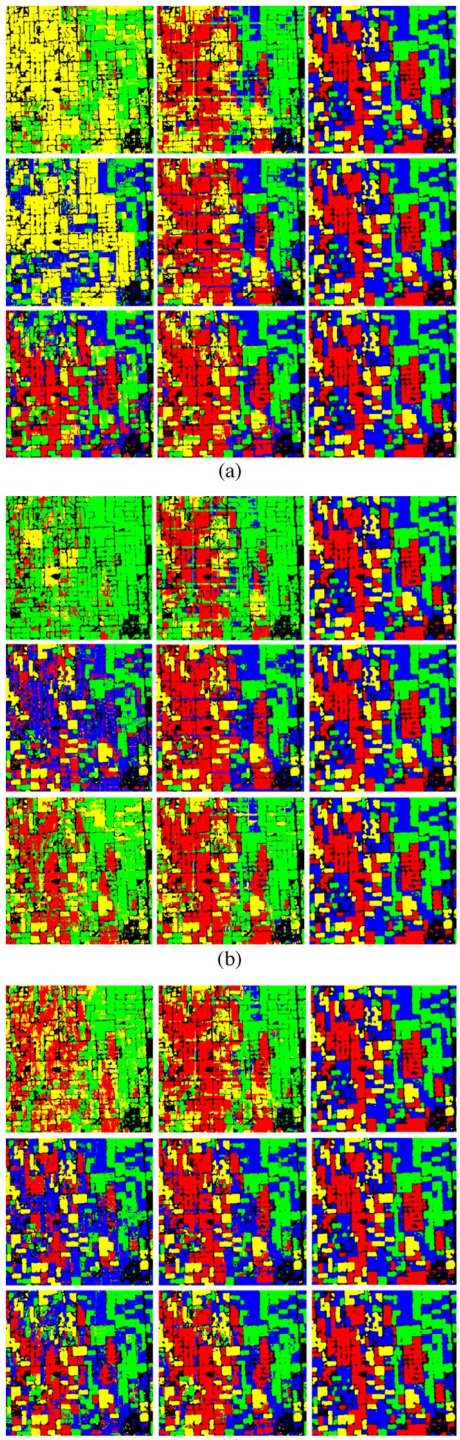

Fig. 4. (a) Crop type maps made with *one* high-resolution image. (A) One high-resolution image early in the year. (B) (A) plus ten low-resolution images. (C) One high-resolution image late in the year. (D) (B) plus ten low-resolution images. (E) One high-resolution image in the middle of the year. (F) (E) plus ten low-resolution images. The color scheme is the same as in Fig. 1. (b) Classification results with two high-resolution images. (A) Two high-resolution images early in the year. (B) (A) plus ten low-resolution images. (C) Two high-resolution images late in the year. (D) (B) plus ten low-resolution images. (E) Two high-resolution image, one in the early and one from late in the year. (F) (E) plus ten low-resolution images. (c) Classification results with three high-resolution images. (A) Three high-resolution images early in the year. (B) (A) plus ten low-resolution images. (C) Three high-resolution images late in the year. (D) (B) plus ten low-resolution images. (E) Three high-resolution image equally spaced out across the year. (F) (E) plus ten low-resolution images. In all cases, the neighborhood heuristic is included. The color scheme is the same as in Fig. 1(c).

TABLE IV
CONFUSION MATRIX FOR ONE HIGH-RESOLUTION IMAGE (OCTOBER 16, 2002) ONLY. THE OVERALL ACCURACY WAS 43.5%

		Ground Truth (Pixels)				
		Corn	Cotton	Rice	Soy	Total
Output Classes (Pixels)	Corn	**17597**	25112	11712	17307	71728
	Cotton	7	**3**	1	17	28
	Rice	1631	1958	**15767**	3550	22906
	Soy	254	9783	2299	**23442**	35778
	Total	19489	36856	29779	44316	**130440**

TABLE V
CONFUSION MATRIX FOR ONE HIGH-RESOLUTION IMAGE (OCTOBER 16, 2002), NEIGHBOR HEURISTIC, AND TEN LOW-RESOLUTION IMAGES. THE OVERALL ACCURACY WAS 62.5%

		Ground Truth (Pixels)				
		Corn	Cotton	Rice	Soy	Total
Output Classes (Pixels)	Corn	**16716**	2932	2087	5410	27145
	Cotton	2193	**26819**	3346	17601	49959
	Rice	460	1792	**20877**	3994	27123
	Soy	120	5313	3469	**17311**	26213
	Total	19489	36856	29779	44316	**130440**

of the rice fields had been incorrectly labeled as corn in the first classification, but these incorrect labels changed to one of the other three crops, with the highest percentage of the pixels changing to rice, the correct label.

Finally, when we look solely at the confusion matrix for the classification with ten low-resolution images, it is clear that a large number of errors remain in distinguishing cotton and soy from one another. Examination of the ideal crop curves shows that the soy curve appears to be most similar to the cotton curve, particularly late in the growing season. This observation clearly demonstrates how the classifier improves when classifying crops with very dissimilar phenologies.

IV. DISCUSSION

Overall, the algorithm performed well on both synthetic and real images. It is encouraging to find that the first ten additional low-resolution images have the greatest impact on classification accuracy, since in many scenarios, it may be difficult to acquire more than ten good low-resolution images for a particular scene.

Regarding improvement when adding more than ten low-resolution images, there were significant differences between synthetic and real data. With synthetic data, accuracy improved an additional 1%–8%, depending on when the high-resolution images were chosen. The greatest gain was when there was only a single high-resolution image taken early in the year. With real data, there was a decrease in accuracy when keeping the low-resolution lambda constant and no change in accuracy from making educated guesses of improved low-resolution lambdas.

One reason for this discrepancy is that the training data used to determine lambda were not adequately representative of the entire scene. Because only one set of training data was used, there is a possibility that the lambda chosen was too small or too large. If we selected a lambda that was too small, the low-resolution data would not have had a strong enough influence on the classification until more low-resolution images were added. If we chose a lambda that was too large, the low-resolution data for the real images would have too strong an influence; thus, as more images were added, their influence would keep increasing and impair the accuracy even more.

A second reason for the lack of increase in overall accuracy using real images with more than ten low-resolution images could be because the weight of the low-resolution images became too large. If more low-resolution images are added while lambda is kept constant, the low-resolution images come to have a higher influence on the outcome of the classifier. The low-resolution term becomes larger as more images are added; thus, eventually, it grows to be more important than the other terms. To compensate for this, the results were recalculated while decreasing lambda as more images were added. An alternate way to compensate would be to normalize each of the terms (high-resolution optimization function, neighbor function, and low-resolution optimization function) so that the lambda values could be kept constant regardless of the number of high- and low-resolution images added.

A third reason for the failure to increase accuracy may have been because the eight-day composited MODIS data were introducing too much noise into the classifier even after choosing the actual compositing days. Due to the way that the eight-day MODIS data are gathered, there are significant noise and spatial uncertainty with each of the images [31].

Regarding improvement using the neighbor heuristic, we found that the change was not as dramatic as the improvement with adding ten low-resolution images. For the synthetic images, we found that the neighbor heuristic improved the classification accuracy between 2% and 8%, which is significant, but still much less than the improvement from the low-resolution images. In the case of the real images, the neighbor heuristic did not improve classification accuracy by any substantial amount. For all experiments, the accuracies with and without the heuristic were nearly identical. Since the truth map for the synthetic images was created under the assumption that it was more likely for neighbors to be of the same class, it may be better to accept the results using real data rather than synthetic data to evaluate the effectiveness of the neighbor heuristic.

The results in map form give a clear picture of how the classification algorithm is behaving and how the results are changed by adding ten low-resolution inputs to the classifier. One of the most obvious changes is that the shapes and sizes of individual fields are much more accurate when using low-resolution inputs and the neighbor heuristic in the classifier. In all of the classification maps generated by using only one high-resolution image, there are many areas where a great majority of the pixels are labeled as one class, yet there are pixels scattered throughout the area that are labeled as other classes. When we consider how the algorithm works, this result is fairly intuitive. If the classifier does not receive any low-resolution inputs or the neighbor heuristic, every pixel is classified autonomously, i.e., without any knowledge of the classification of its neighbors. However, when the neighbor function and the low-resolution inputs are added, these pixels are now influenced by their neighbors, and they also benefit from information about the larger area, owing to the low-resolution pixels that present an average over a larger space. Of most importance is the temporal information low-resolution images provide to help detect what classes are most likely to be found in a given spatial area. This

influences the pixels that are hard to classify, pushing them in the right direction, with noticeably improved results.

Note that, in the maps, there are occasional systematic horizontal lines of incorrectly labeled crops. This problem occurs because of the way the original image was split up and reassembled. Because the Landsat and MODIS data did not overlap perfectly, the four top Landsat pixels for each of the smaller images were outside the range of the MODIS pixels; thus, they did not receive the same bias from the low-resolution images as the rest of the pixels. To remedy this problem in future runs, the original Landsat images must be resampled so that they better overlap the MODIS image grid layout.

Several generalizations about the particularly useful applications of the algorithm can be made. First, the minimum requirements for the algorithm include one or more high-resolution images and more than one low-resolution image of the same scene, the presence of similar spectral bands in both high- and low-resolution images, and ideal crop reflectance curves for all of the crops in question. If these ideal crop reflectance curves are not available from field measurements, they can be generated from a time series image, perhaps from a different time and location. They can also be generated theoretically as well as with the help of crop growth models.

Our results show that, if only one or two high-resolution inputs and ten or more low-resolution inputs are available, this approach is likely to improve classification accuracy. On the other hand, if three or more high-resolution inputs are available, then our approach may not significantly improve the classification map. In the trials, we found that, with three images, classification accuracy is likely to increase by 10% or less.

The time of the year when the high-resolution images are acquired also influences the level improvement with the proposed method. If they are taken at times when the reflectance values of the crops are similar, then the addition of low-resolution inputs to the classifier will be more likely to improve accuracy. To test for the similarity of reflectance values, the ideal crop reflectance curves can be used. If three or more high-resolution images are available, and these images had been made at times of the year when there are large differences in the reflectance values of the crops under study, then our approach may not be needed. We show that, when the high-resolution images were all from the beginning of the year, when reflectance values of the four crops were most similar, the accuracy in the absence of low-resolution images was very poor, and addition of low-resolution images boosted the classification accuracy by 20% in numerous instances. On the other hand, if the single high-resolution image was from the middle of the season, the accuracy of the crop type map without the low-resolution input would be considerably better.

When the user has access to only one or two high-resolution images but with unlimited number of low-resolution data, it would be wise to look first at the ideal crop curves and choose times of the year when the differences in reflectance values are greatest for that set of crops. The use of a classifier such as ours provides an additional tool to assess and diminish the tradeoff between classification accuracy and the cost of purchasing high-resolution images. We show that, if access to a broad array of low-resolution images is possible and only one high-resolution image is available, classification accuracy can be improved to almost the same level as if there were two high-resolution images. Furthermore, if two high-resolution images and numerous low-resolution images are utilized, the classification accuracy can be boosted as if three high-resolution images were available. Thus, it appears that the set of low-resolution images improves the classification accuracy by about the same amount as the addition of another high-resolution image would. This improvement is particularly useful given that the availability of high-resolution imagery is often limited due to acquisition strategies, cloud presence, and/or cost.

One limitation of the present algorithm is its speed. It was implemented in Python, a high-level programming language that is significantly slower than traditional applications using FORTRAN or other low-level languages. Furthermore, the choice of algorithm also severely limits speed. While the PGD is effective in finding the correct answer, there are major issues involving speed and memory usage. There are a number of locations where speed gains are possible. First is the search step of PGD. It is possible to decrease the interval of search by half or even a quarter (i.e., instead of searching every instance, we can search every other or every fourth instance). While we have not implemented this operation, we estimate better than twofold increase in speed by halving the search step. With a preprocessing step, it may also be possible to locate potential start and end points within the search space and do more intensive search within that window instead of searching the entire parameter space. It may also be possible to incorporate the timing of high-resolution image into the search algorithm as a preprocessing. There is an obvious relationship between the timing of the high-resolution image during the growing season and the speed with which the solution is found in the optimization process. Alternately, it may be worthwhile to investigate a stochastic gradient descent method as a solution. This approach has the advantage that the full gradient does not need to be computed at every iteration but rather randomly, and thus, not all data need to reside in memory at all times.

As with all classification problems using remote sensing, one limitation of this work may be one-place one-time syndrome in which the findings from this study are not easily generalized to other and larger areas. We have reasons to believe that our study can indeed be applicable across large areas and other locations around the world, as long as input data sets are available, and computer resources can be devoted to process images. For example, we tested our work in the Yaqui Valley in western Mexico (not shown) and successfully mapped two out of three crop types, using similar sets of inputs. Moreover, the work presented here considers Landsat-like observations as high-resolution while MODIS data as low-resolution images, which is directly related to the size of the agricultural fields in the study area. There are many places around the world, such as China and Africa, where even Landsat-like observations may be considered low-resolution simply due to small field sizes [12]. In these situations, it is possible to use at least one very high resolution (< 5 m) data such as those provided by many private sensors and incorporate more frequent Landsat data (if available) as the low-resolution input. While this approach is

not tested here, the findings of our study suggest promising outcomes in such applications. The most limiting parts of that work would be the following: 1) matching of spectral response across bands and 2) finding sufficient Landsat-like observations throughout the growing season.

In addition to improved accuracy, another benefit of our classification approach over other crop type mapping exercises is that, once ideal crop curves are generated, the classification algorithm can be used in any agricultural landscape, given that the crop types for which the ideal crop curves exist are present on the landscape. With the new algorithm, the user relies on ideal crop curves rather than labels and does not need to spend time labeling training instances. For instance, the user with access to ideal crop curves could potentially construct an application that would programmatically classify any given area where those crop curves were relevant. Given the degree of automation and abstraction occurring in today's remote sensing world, this would make it that much easier to map crop types than the traditional methods.

V. Conclusion

The goal of this paper was to present a new approach to crop type classification that uses information from images of different spatial and temporal resolutions simultaneously, by exploiting the strengths of both. Testing the algorithm with both synthetic and real images showed that merging high-resolution images with low-resolution data produced better crop type maps than when using high-resolution images alone. If only the high-resolution data are available, the results are most accurate when the high-resolution images are as follows: 1) more than one; 2) acquired during the year when crops are most distinguishable; and 3) include at least one peak growing season image. Adding the low-resolution data results in improving the accuracies partly because adding roughly ten low-resolution images functions as adding single-high-resolution-image data to the classification process. At the same time, the error structure that results from only high-resolution inputs is maintained—if there is good temporal placement of high-resolution data, the overall accuracies are higher and get marginally boosted with the addition of at least ten low-resolution images.

While not tested extensively in other locations or with different inputs, the proposed method has the potential to improve crop type mapping in many places around the world, particularly in locations where crop fields are small. The most limiting parts of such application would be the following: 1) matching of spectral response across bands and 2) finding sufficient Landsat-like observations throughout the growing season.

Detailed and accurate crop type maps are important for many reasons, and it is an ongoing effort for the remote sensing community to continue to develop methods for producing improved crop maps. There is a plethora of readily available remotely sensed data, and it is crucial that our community continues to develop new ways to extract ever more useful information from these. The approach presented here, merging two different types of remotely sensed images, attempts to broaden the scope of what is possible within the realm of remote sensing.

Acknowledgment

The authors would like to thank Mr. G. Allez and Dr. A. Schneider for their careful editing of an earlier draft of this manuscript that improved the clarity and the readability of the manuscript. They would also like to thank three anonymous reviewers whose suggestions greatly improved the contents of this paper.

References

[1] J. de Leeuw, Y. Georgiadou, N. Kerle, A. de Gier, Y. Inoue, J. Ferwerda, M. Smies, and D. Narantuya, "The function of remote sensing in support of environmental policy," *Remote Sens.*, vol. 2, no. 7, pp. 1731–1750, Jul. 2010.

[2] P. F. Crapper, "Errors incurred in estimating an area of uniform land cover using Landsat," *Photogramm. Eng. Remote Sens.*, vol. 46, no. 10, pp. 1295–1301, 1980.

[3] G. D. Badhwar, "Automatic corn soybean classification using Landsat MSS data: Early season crop proportion estimation," *Remote Sens. Environ.*, vol. 14, no. 1–3, pp. 31–37, Jan. 1984.

[4] F. Crook, "Underreporting of China's cultivated land area: Implications for world agricultural trade," US Department of Agriculture, Washington, DC, USA, China International Agricultural and Trade Rep. No. RS-93-4, 1993, Situation and Outlook Series (33–39).

[5] P. S. Thenkabail, P. Gangadhara Rao, T. Biggs, M. Krishna, and H. Turral, "Spectral matching techniques to determine historical land use/land cover (LULC) and irrigated areas using time-series AVHRR pathfinder datasets in the Krishna River Basin, India," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 9, pp. 1029–1040, Sep. 2007.

[6] J. H. Smith and D. B. Ramey, "A crop area estimator based on changes in the temporal profile of a vegetative index," in *Proc. Survey Res. Methods Section*, 1982, pp. 495–498.

[7] F. G. Hall and G. D. Badhwar, "Signature-extendable technology: Global space-based crop recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-25, no. 1, pp. 93–103, Jan. 1987.

[8] K. P. Price, S. L. Egbert, M. S. Nellis, R. Y. Lee, and R. Boyce, "Mapping land cover in a high plains agro-ecosystem using a multidate Landsat thematic mapper modeling approach," *Trans. Kansas Acad. Sci.*, vol. 100, no. 1/2, pp. 21–33, Apr. 1994.

[9] P. S. Thenkabail, R. B. Smith, and E. De-Pauw, "Evaluation of narrow-band and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization," *Photogramm. Eng. Remote Sens.*, vol. 68, no. 6, pp. 607–621, Jun. 2002.

[10] B. D. Wardlow, S. L. Egbert, and J. H. Kastens, "Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great plains," *Remote Sens. Environ.*, vol. 108, no. 3, pp. 290–310, Jun. 2007.

[11] M. Ozdogan, "The spatial distribution of crop types from MODIS data: Temporal unmixing using independent component analysis," *Remote Sens. Environ.*, vol. 114, no. 6, pp. 1190–1204, Jun. 2010.

[12] M. Ozdogan and C. E. Woodcock, "Resolution dependent errors in remote sensing of cultivated areas," *Remote Sens. Environ.*, vol. 103, no. 2, pp. 203–217, Jul. 2006.

[13] S. P. Vyas, M. P. Oza, and V. K. Dadhwal, "Multi-crop separability study of Rabi crops using multi-temporal satellite data," *J. Indian Soc. Remote Sens.*, vol. 33, no. 1, pp. 75–79, Mar. 2005.

[14] D. M. Johnson and R. Mueller, "The 2009 cropland data layer," *Photogramm. Eng. Remote Sens.*, vol. 76, no. 11, pp. 1201–1205, 2010.

[15] C. Boryan, Z. Yang, R. Mueller, and M. Craig, "Monitoring US agriculture: The US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program," *Geocarto Int.*, vol. 26, no. 5, pp. 341–358, 2011.

[16] J. B. Adams, M. O. Smith, and P. E. Johnson, "Spectral mixture modeling: A new analysis of rock and soil types at Viking Lander site," *J. Geophys. Res.*, vol. 91, no. B8, pp. 8098–8112, Jul. 1986.

[17] N. A. Quarmby, J. R. G. Townshend, J. Settle, K. H. White, M. Milnes, T. L. Hindle, and N. Silleos, "Linear mixture modeling applied to AVHRR data for crop area estimation," *Int. J. Remote Sens.*, vol. 13, no. 3, pp. 415–425, Feb. 1992.

[18] D. B. Lobell and G. P. Asner, "Cropland distributions from temporal unmixing of MODIS data," *Remote Sensing of Environment*, vol. 93, no. 3, pp. 412–422, Nov. 2004.

[19] P. S. Thenkabail, C. M. Biradar, P. Noojipady, V. Dheeravath, Y. J. Li, M. Velpuri, M. Gumma, G. P. O. Reddy, H. Turral, X. L. Cai, J. Vithanage, M. Schull, and R. Dutta, "Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium," *Int. J. Remote Sens.*, vol. 30, no. 14, pp. 3679–3733, 2009.
[20] X. Xiao, J. Liu, D. Zhuang, S. Frolking, S. Boles, B. Xu, M. Liu, W. Salas, B. Moore, and C. Li, "Uncertainties in estimates of cropland area in China: A comparison between an AVHRR-derived dataset and a Landsat TM-derived dataset," *Global Planet. Change*, vol. 37, no. 3/4, pp. 297–306, Jun. 2003.
[21] M. Hansen, D. P. Roy, E. Lindquist, C. Justice, and A. Alstatt, "A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in Central Africa," *Remote Sens. Environ.*, vol. 112, pp. 2495–2513, 2008.
[22] A. Fischer, "A simple model for the temporal variations of NDVI at regional scale over agricultural countries: Validation with ground radiometric measurements," *Int. J. Remote Sens.*, vol. 15, no. 7, pp. 1421–1446, May 1994.
[23] C. J. Kucharik, "Evaluation of a process-based agro-ecosystem model (AGRO-IBIS) across the U.S. Cornbelt: Simulations of the inter-annual variability in maize yield," *Earth Interactions*, vol. 7, no. 14, pp. 1–33, Dec. 2003.
[24] C. Huang, J. R. G. Townshend, S. Liang, S. N. V. Kalluri, and R. S. DeFries, "Impact of sensor's point spread function on land cover characterization: Assessment and deconvolution," *Remove Sensing of Environment*, vol. 80, no. 2, pp. 203–212, 2002.
[25] N. Z. Shor, *Minimization Methods for Non-differentiable Functions*. New York, NY, USA: Springer-Verlag, 1985.
[26] C. O. Justice, E. Vermote, J. R. G. Townshend, R. DeFries, D. P. Roy, D. P. Hall, V. V. Salomonson, J. L. Privette, G. Riggs, A. Strahler, W. Lucht, R. Myneni, Y. Knyazikhin, S. W. Running, R. R. Nemani, Z. Wan, A. Huete, W. van Leeuwen, R. E. Wolfe, L. Giglio, J.-P. Muller, P. Lewis, and M. J. Barnsley, "The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 4, pp. 1228–1249, Jul. 1998.
[27] C. B. Schaaf, F. Gao, A. H. Strahler, W. Lucht, X. Li, T. Tsang, N. C. Strugnell, X. Zhang, Y. Jin, J.-P. Muller, P. Lewis, M. Barnsley, P. Hobson, M. Disney, G. Roberts, M. Dunderdale, C. Doll, R. P. d'Entremont, B. Hu, S. Liang, J. L. Privette, and D. P. Roy, "First operational BRDF, albedo nadir reflectance products from MODIS," *Remote Sens. Environ.*, vol. 83, pp. 135–148, 2002.
[28] J. G. Masek, C. Huang, R. Wolfe, W. Cohen, F. Hall, J. Kutler, and P. Nelson, "North American forest disturbance mapped from a decadal Landsat record," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 2914–2926, Jun. 2008.
[29] N. Guindin-Garcia, A. A. Gitelson, T. J. Arkebauer, J. Shanahan, and A. Weiss, "An evaluation of MODIS 8- and 16-day composite products for monitoring maize green leaf area index," *Agricultural Forest Meteorol.*, vol. 161, pp. 15–25, Aug. 2012.
[30] C. E. Woodcock and V. J. Harward, "Nested-hierarchical scene models and image segmentation," *Int. J. Remote Sens.*, vol. 13, no. 16, pp. 3167–3187, Dec. 1992.
[31] B. Tan, C. E. Woodcock, J. Hu, P. Zhang, M. Ozdogan, D. Huang, W. Wang, Y. Knyazikhin, and R. B. Myneni, "The impact of gridding artifacts on the local spatial properties of MODIS data: Implications for validation, compositing, and band-to-band registration across resolutions," *Remote Sens. Environ.*, vol. 105, no. 2, pp. 98–114, Nov. 2006.

Mark W. Liu received the B.S. degree in computer science and engineering from the University of California, Los Angeles, CA, USA, in 2007 and the M.S. degree in computer science and the M.S. degree in environment and resources from the University of Wisconsin, Madison, WI, USA, in 2009 and 2011, respectively.

He is currently the Vice-President of Engineering with 8 to 18 Media, Chicago, IL, USA. He was the founder of a sports statistics company called Leaguevine that was subsequently acquired by 8 to 18 Media. His present role revolves around software engineering, ranging from product vision to implementation.

Mutlu Ozdogan was born in Turkey in 1971. He received the B.S. degree in earth sciences from Istanbul University, Istanbul, Turkey, in 1995, the M.S. degree from North Carolina State University, Raleigh, NC, USA, and the M.S. degree in environmental studies and the Ph.D. degree in geography and environment (in 2005) from Boston University, Boston, MA, USA.

From 2005 to 2007, he was a Postdoctoral Scientist with the Goddard Space Flight Center, National Aeronautics and Space Administration, Greenbelt, MD, USA, under funding from the National Research Council. Since then, he has been an Assistant Professor with the Department of Forest and Wildlife Ecology and the Nelson Institute for Environmental Studies, University of Wisconsin, Madison, WI, USA. He is the author of more than 30 papers. His research interests include remote sensing of crops and crop yields, land-use changes, hydrology, and water resources.

Xiaojin (Jerry) Zhu received the B.S. and M.S. degrees in computer science from Shanghai Jiao Tong University, Shanghai, China, in 1993 and 1996, respectively, and the Ph.D. degree in language technologies from Carnegie Mellon University, Pittsburgh, PA, USA, in 2005.

From 1996 to 1998, he was a Research Staff Member with IBM China Research Laboratory, Beijing, China. Since 2005, he has been a Professor with the Department of Computer Sciences, University of Wisconsin, Madison, WI, USA. His research interest is in machine learning.

Dr. Zhu was a recipient of awards and honors, including a National Science Foundation CAREER Award and paper awards at International Conference on Machine Learning (ICML), The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD), and Special Interest Group on Software Engineering (SIGSOFT).