Boundary Delineation of Agricultural Fields in Multitemporal Satellite Imagery

Heather C. North, *Member, IEEE*, David Pairman, *Member, IEEE*, and Stella E. Belliss

***Abstract*—Agricultural land-use statistics are more informative per-field than per-pixel. Land-use classification requires up-to-date field boundary maps potentially covering large areas containing thousands of farms. This kind of map is usually difficult to obtain. We have developed a new, automated method for deriving closed polygons around fields from time-series satellite imagery. We have been using this method operationally in New Zealand to map whole districts using imagery from several satellite sensors, with little need to vary parameters. Our method looks for boundaries—either step edges or linear features—surrounding regions of low variability throughout the time series. Local standard deviations from all image dates are combined, and the result is convolved with a series of extended directional edge filters. We propose that edge linearity over a long distance is a more important criterion than spectral difference for separating fields, so edge responses are thresholded primarily by length rather than strength. The resulting raster edge map (combined from all directions) is converted to vector (GIS) format and the final polygon topology is built. The method successfully segments parcels containing different crops and pasture, as well as those separated by boundaries such as roads and hedgerows. Here we describe the technique and demonstrate it for an agricultural study site (4000 km^2) using SPOT satellite imagery. We show that our result compares favorably with that from existing segmentation methods in terms of both quantitative quality metrics and suitability for land-use classification.**

***Index Terms*—Agriculture, feature extraction, image edge analysis, image segmentation, image sequence analysis, remote sensing.**

I. Introduction

REMOTE sensing imagery can provide detailed, up-to-date, and spatially explicit information on agricultural land use that would otherwise be difficult to gather. Field surveys and farmer questionnaires are time consuming and have their limitations for updating large areas. If remotely sensed imagery is interpreted manually, this is also labor intensive, so many researchers have worked on automated techniques for classifying crops and land uses, often using time series of images; e.g., Hall and Badhwar [1], Lucas *et al.* [2], Wardlow *et al.* [3], and Esch *et al.* [4]. Our own work in agricultural land-use mapping is described in Lilburne and North [5], North *et al.* [6], [7], and in a technical report [8]. Such classification is increasingly carried out on a per-field rather than a per-pixel basis [2], [4], [8]. Agricultural statistics are more useful in this form for policy and monitoring, and classification techniques are more powerful and accurate when whole fields are classified as a unit [9], since, as an object, a field can be robustly characterized by its mean reflectance, and higher level features such as size, shape, and texture can also be considered.

Field boundary information is often available for individual farms, but it is difficult to collate and maintain this in a uniform way across a region containing thousands of farms. Also, physical fence-lines are not the only feature of interest: sometimes several different crops are grown inside a single field. Since the final aim is to separately classify these parcels of crop, subareas must also be segmented. Rydberg and Borgefors [10] define field boundaries as being at locations "where a change in crop type takes place or where two similar crops are separated by a natural disruption in the landscape, like a ditch or a road." To this we add any significant differences in crop management.

To be practical, any boundary mapping method must be able to work across large areas, such as whole districts or full satellite scenes. It must also be easy to quickly regenerate the field boundary map from new image sequences so that the result can be kept up to date. Changes in farm type often give rise to new field layouts, and subareas of crop within fields vary from year to year.

Existing approaches for segmenting imagery include:

1) edge detection methods [11], [12]—these accurately locate significant boundaries but do not guarantee closed polygons;
2) region-based methods, involving progressively merging adjacent areas that have similar spectral properties (bottom up, region growing) [13], [14]; or splitting areas that have different spectral properties (top down) [15]; or clustering spectrally similar pixels around a set of *k*-means [16]—all produce closed polygons, but the boundaries are not always located at the natural/visible edges of the highest gradient or linearity;
3) integrated methods that seek to combine the advantages of both edge and spectral approaches [10], [17] and, more recently, [18] and [19].

These techniques can be further combined with other metrics on potential object segments, such as shape and size. Commercial toolkits such as eCognition [20] are available for developing segmentation and object classification algorithms [21], [22]. The

Manuscript received August 20, 2017; revised February 12, 2018 and September 24, 2018; accepted November 19, 2018. This work was supported by Manaaki Whenua—Landcare Research Strategic Science Investment Fund from the Ministry of Business, Innovation and Employment (MBIE). *(Corresponding author: Heather North.)*

The authors are with Manaaki Whenua—Landcare Research, PO Box 69040, Lincoln 7640, New Zealand (e-mail: northh@landcareresearch.co.nz; pairmand@landcareresearch.co.nz; BellissS@landcareresearch.co.nz).

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/JSTARS.2018.2884513

precise algorithm used depends on the image type, its content, and which features are of interest.

The development of our method (our early work is described in [23]) has been guided by our observation of several key characteristics of the agricultural landscape, as seen in the imagery.

1) Many crops and pastures are very similar in spectral appearance when they are at full leaf cover, so segmentation by spectral difference alone is not always sufficient. For the same reason, edges between fields can sometimes be extremely weak (particularly those between neighboring pasture fields), meaning they are not always picked up by a simple edge detector operating on edge strength.
2) Agricultural fields often have straight boundaries, or, if curved, then the curve is usually gentle. We therefore consider linearity over a long distance (multiple pixels) to be a strong indicator that an edge response is due to a fence-line or change in crop type or management, even if the edge is very weak.
3) A field, or area of crop, suitable for classification as a unit is, by definition, internally homogeneous. We aim to find boundaries around these internally consistent areas. We use standard deviation (STD) within a moving window in each image band or channel. A number of authors have used elevated STD, variance, or coefficient of variation to indicate the presence of boundaries, including both linear features and step edges (stepping either up or down); e.g., [24]–[26].
4) Crop rotation, pasture management, and crop phenology through a season can cause large variation in the appearance of fields through time (more so for arable cropping fields than for pasture fields). We exploit this by using time series of images: even if neighboring fields are indistinguishable in one image, there is a good chance they will look different in another, and therefore an edge will be visible between them. Per-band local STD is combined at each pixel through all bands of all images in the time series.

Most existing methods for edge detection and segmentation of objects in satellite imagery are based on the magnitude of spectral difference from one object to another. As noted, we believe this is not necessarily the most useful criterion for separating agricultural fields. For example, we do not wish to merge two field segments on the basis of spectral similarity if they are separated by a linear feature such as a hedgerow.

An important case of weak edges are those between neighboring pasture fields. While some authors focus on delineation of arable or harvested crops [19], [27], it is important in the New Zealand context to also delineate pasture fields. Most regions include a mix of arable and pastoral farming. Pasture grazing is often managed intensively, especially in dairy farming. Neighboring pasture fields are often separated only by a wire fence, with no visible linear feature in the image. The indicator of boundary presence in this case is a difference in pasture length (e.g., pre- or post-grazing) that may appear in some images of the time series.

Therefore, we have developed a method that places its main emphasis on the length and linearity of boundary segments. Rydberg and Borgefors [10] and Yan and Roy [27] also aim to retain linear edges, even where these are weak. However, as described in the following sections, our method takes this concept further by using extended directional linear filters for edge detection, and by thresholding primarily using length, rather than strength, of edge responses. Our method converts this initial raster line work to closed GIS polygons, a step suggested by Rydberg and Borgefors [10] as a logical extension of their own work.

The use and combination of directional filters and the thresholding of their responses using length rather than strength is the key contribution of this work. Other features of our method are as follows.

1) The perimeters of individual image extents and data gaps (e.g., from clouds) do not create artifacts—our method can combine data from a time series of incomplete images, so gap-filling methods such as those described in [28] or [19] are not required.
2) It is fully automated, with few parameters, which are not generally changed between datasets—we typically use medium resolution imagery from the SPOT, Sentinel-2, and occasionally Landsat satellites (sometimes in combination) without changing parameters.
3) It is operationally practical for large areas because imagery is processed as overlapping tiles.

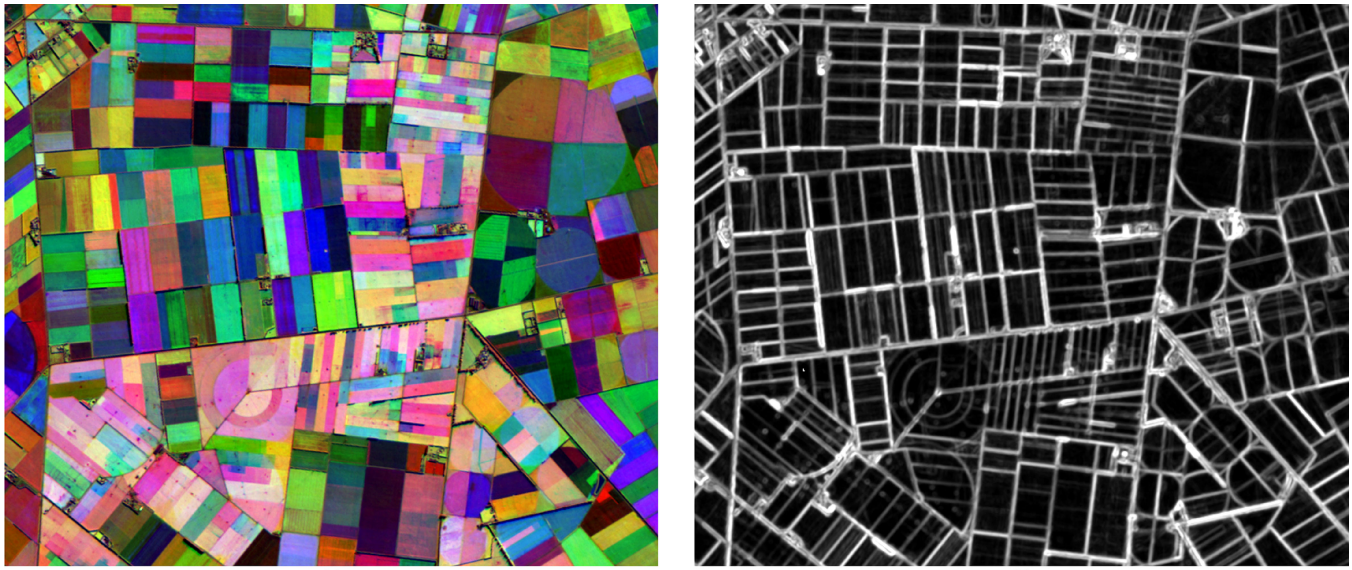
In the following sections we list the steps in our new method, consisting first of processing in raster space, followed by conversion to a final GIS polygon layer—a representation of fields as objects. We describe the execution of the method on a test dataset of time-series satellite imagery, and present a subarea of the resulting field boundary map. We then compare the map quantitatively and descriptively with results from another method (the Erdas Imagine image segmentation tool [29]) operated on the same image dataset. We also show comparison results with a second existing method [30], and the operation of our own method with the same parameters on three other datasets, using a variety of image types, to demonstrate its robustness.

II. Method

A. Field Boundary Detection Method—Raster Analysis

Our method operates on a time series of satellite images to increase the possibility that any given boundary will be visible in at least one of the images. The methodology below is illustrated with figures that show a 6 km $\times$ 7 km subset [see Fig. 1(a)] of the full test dataset described and used in Section III. For now it is sufficient to note that the red, near-infrared (NIR), and short-wave infrared (SWIR) bands from seven dates of SPOT satellite imagery were used to generate these figures. We often use this combination of bands because they all show useful differences between fields and are relatively uncorrelated. However, it is quite possible to use, for example, just the NIR band, or an index such as the Normalized Difference Vegetation Index (NDVI), or band ratios (particularly for hill country), or principal components images.

The detailed steps for generating the STD image and detecting long, linear features are given below.



(a) (b)

Fig. 1. (a) NIR bands of three image dates, displayed as red–green–blue (images CNES 2010 and 2011), to show the general layout of fields. (b) Resulting image of combined STDs for seven images using three bands each.

1) Before calculating the per-band STD, a weight can be applied to each of the bands to adjust their influence with respect to one another. For the normal cases where image bands are calibrated to reflectance or used as ratios of reflectance, they are already balanced, and we use a weighting of 1.0 for all bands. However, if the imagery is calibrated to a different unit, e.g., radiance, or there is a reason to emphasize one band over another, then weighting parameters other than 1.0 can be used. In such cases, we define a set of weightings suitable for a class of image (i.e., combination of sensor, calibration method, and bands used) rather than for individual images. We consider one of the bands as a reference (weighting of 1.0), and calculate the ratio of the average dynamic range (over a large number of images) of this band to each other one. If each image has n bands, b_k where $k = 1, \ldots, n$, and the weights are w_k then each band is weighted as $w_k . b_k$.
2) For each image (and each weighted band), the STD is calculated at every point using the pixels in a local area around that point. We use a round window of diameter 5 pixels to define this local area (i.e., 21 pixels). Invalid pixels (e.g., outside the satellite's path, cloud-contaminated, or other gaps) do not contribute to the STD. This step results in a stack of per-band STD images.
3) The STDs from all images and all bands are combined using an average at each point to obtain a single-layer image summarizing spatial variability from all images and all bands. Again, only STDs from valid imagery are combined at this point. An example of a combined STD image S is shown in Fig. 1(b).
4) Linear features in the combined STD image are found by convolution with 16 pairs of directional operators, where each pair consists of a left and a right half, $f_{i,\text{left}}$ and $f_{i,\text{right}}$ for $i = 1, \ldots, 16$ directions. Consecutive operators are 11.25° apart, so the 16 together occupy 180°. Fig. 2 shows the first five of the operator pairs: the remaining 11 can be derived from these by rotation and reflection. All have a length equivalent to 13 pixels. Processing with these produces a 16-layer image B with each layer i representing a different direction.

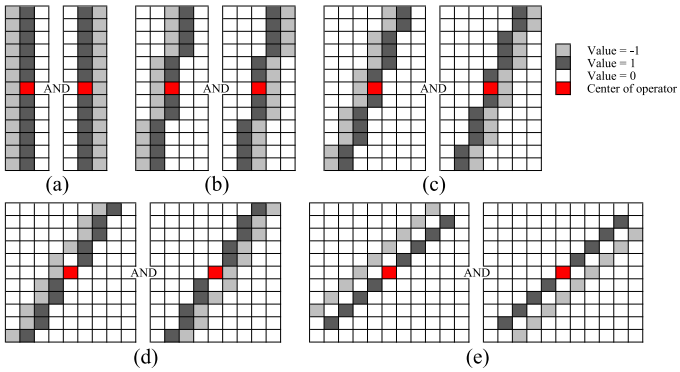


Fig. 2. Five of the 16 directional convolution operator pairs, covering the first 45° of the full 180°. (a) 0°. (b) 11.25°. (c) 22.5°. (d) 33.75°. (e) 45°.

5) Local maxima lines in the combined STD image are found using a logical (AND) combination of the directional operator pairs. Operator responses are

$$L_{i,\text{left}} = f_{i,\text{left}} \otimes S \tag{1a}$$

$$L_{i,\text{right}} = f_{i,\text{right}} \otimes S. \tag{1b}$$

A boundary response meets the following criteria. At each pixel of B_i

$$B_i(x,y) = \begin{cases} 1, & L_{i,\text{left}}(x,y) > 0,\ L_{i,\text{right}}(x,y) > 0, \\ & (L_{i,\text{left}}(x,y) + L_{i,\text{right}}(x,y)) > T \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where $(x,\ y)$ is a spatial pixel location.

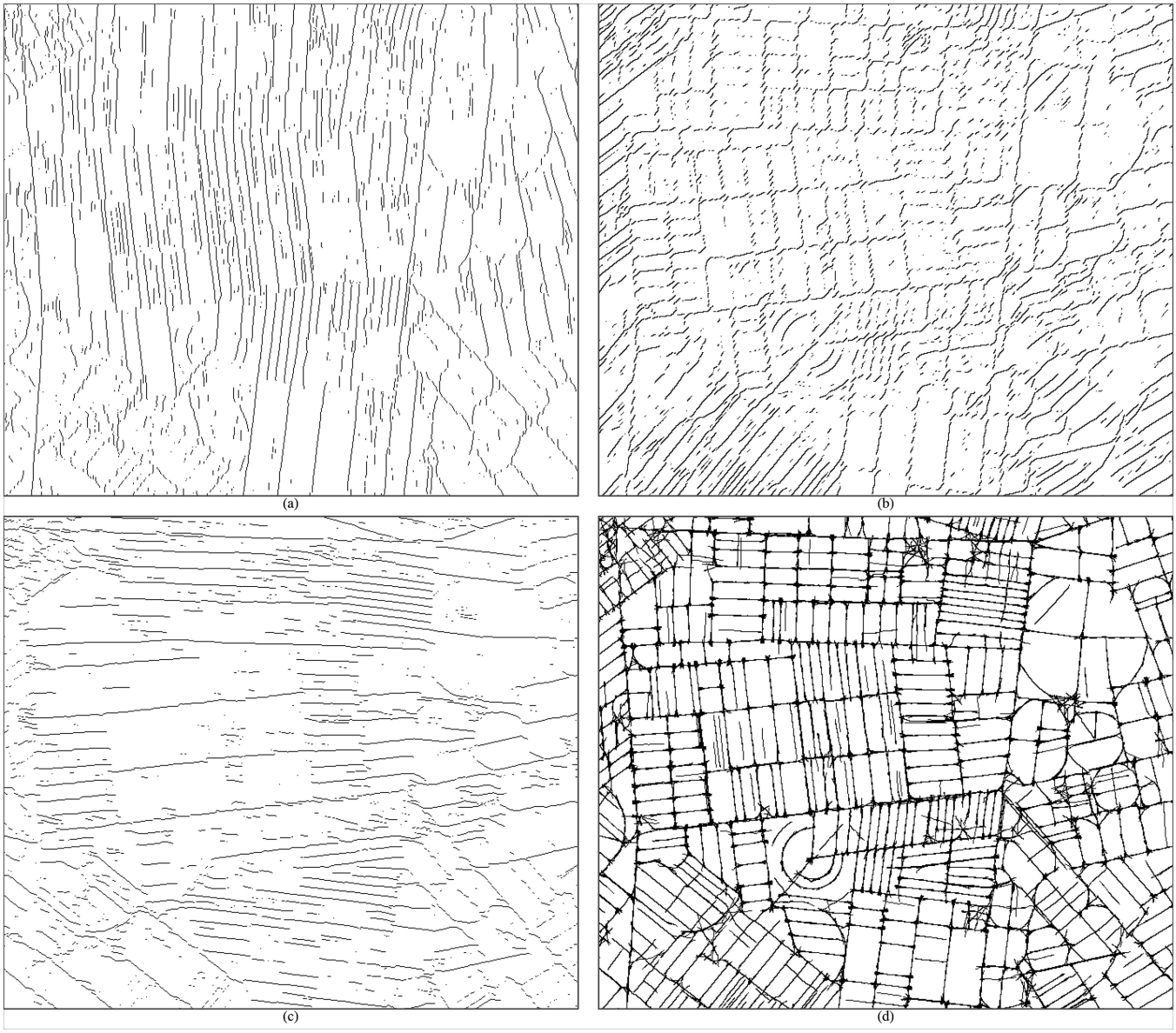


Fig. 3. Responses from three of the 16 directional operators processed by (2), and prior to sieveing by length, showing (a) 0°, (b) 45°, and (c) 100° (101.25)°; and (d) the final raster line work after connectivity analysis and removal of short fragments in each directional response, and then combining all 16 into a single layer.

Only an extremely low threshold T is used on the operator responses to screen out very low-level noise. It is set at a level aimed to retain any edge response that could possibly be significant, as discussed further below. An example of the binary result from the responses to three of the 16 directional operator pairs processed by (2) is shown in Fig. 3.

6) In each of the 16 layers, a connected components analysis of the pixels is done using an eight-neighborhood to label connected line fragments.
7) Ideally, fragments with a length less than approximately 300 m are removed. We use a simple count of pixels belonging to the fragment. Thus, for 10 m SPOT imagery, we remove (sieve) fragments of less than 30 pixels. The 300 m threshold was chosen as appropriate to the usual field sizes and layouts in New Zealand.
8) All 16 angles are combined into a single layer (binary) using a logical OR [see Fig. 3(d)]. The line work is highly connected once all the angles are combined.

Even though our raster analysis algorithm is essentially an edge detection method, it uses a characteristic more often associated with region-based methods: that of spectral uniformity, via the local STD. We emphasize and extend long linear features by using long, narrow operators. This helps us to achieve the highly connected line work result. For agricultural fields, we have observed that it is primarily the edge length in a given direction that indicates an edge response is significant, rather than its strength. Therefore, we choose not to use the hysteresis thresholding approach of Canny [11] because it is not necessary for an edge fragment to contain pixels of high strength (e.g., above the upper threshold of [11])

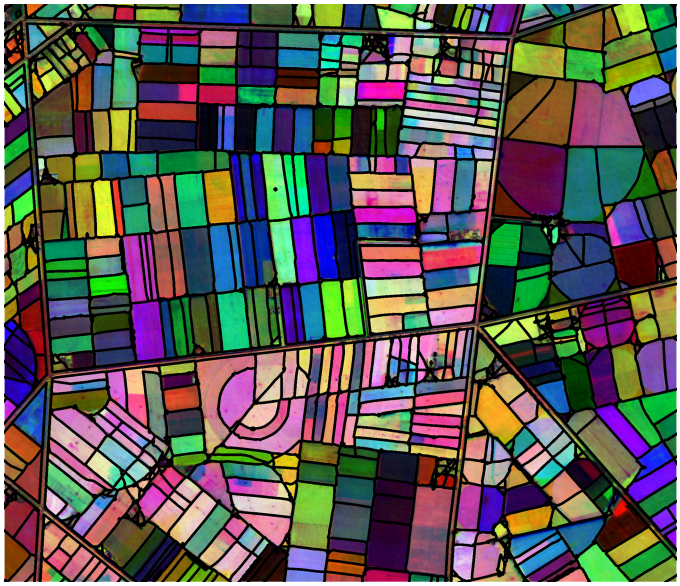


Fig. 4. Final vector line work for the same area shown in Figs. 1 and 3, laid over three-date NIR image.

for it to be considered evidence of a fence-line or change in crop.

We use only an extremely low-strength threshold, set at 10% of the 95th percentile of edge strengths, to screen out the very lowest background noise within fields. This is because we wish to retain any long linear edge, no matter how weak. Yan and Roy [27] also use "aggressively low thresholds to ensure weak edges are not missed . . . " They determined these by manual inspection of the edge intensity values of weak field boundaries. Our length thresholding method, applied to each edge direction separately, is a key feature in achieving the high connectivity, result seen in Fig. 3(d).

The operator pairs do not respond to gradients in STD (i.e., positive response on one side and negative on the other) because we calculate each side separately and require both to be positive. This detects the maxima lines and so keeps the operator response narrower than the full width of the STD response. Also, there are some cases where there is more variability in one part of a field than in another, without an actual boundary between them, and the operator does not respond to this change in variability.

B. Field Boundary Detection Method—GIS Analysis

Raster line work from the steps above [see Fig. 3(d)] provides mostly complete boundaries between fields, but some gaps remain. Before generating a GIS layer with closed polygon topology defining fields, we need to bridge these gaps, where appropriate, and discard other spurious "dangles" (lines that do not connect at both ends). Therefore, the next stage is to run a series of operations (including several heuristics) using the open source GRASS GIS [31] to clean the raster result [see Fig. 3(d)] and convert it into the final vector polygon coverage shown in Fig. 4.

1) Raster lines are thinned to a single pixel using a standard GRASS function (r.thin), an implementation of [32].
2) Urban areas, roads, river beds, lakes, forests, mountainous, and other nonagricultural areas are masked out in raster space using a separately prepared "agricultural mask." For this mask we use nationally available GIS layers defining conservation estate, road networks, and broad land covers, supplemented by manual checking around marginal areas such as river banks. Government cadastral data are also used by Yan and Roy [27] for separating out agricultural land, while in [18] they detect areas with consistently high NDVI, and Graesser and Ramankutty [19] use semisupervised classification.
3) Thinned lines are converted into a GIS vector layer.
4) Lines are smoothed to remove staircasing using the GRASS "snakes" line generalization method described in [31], with default parameters.
5) The agricultural mask vector boundaries are added. Note that this includes roads (buffered by 30 m).
6) The line work is then cleaned (GRASS v.clean) to remove duplicates and short dangles (under 1.5 pixels) to reduce the processing load of step 7.
7) Longer dangles are extended in their current direction by a distance up to the lesser of 50% of their original length or 200 m, to allow intersection with the first encountered line or other extended dangle. We have found these heuristics ensure sufficient support from existing line work to extend across most gaps that should be bridged, without introducing spurious boundaries by extending short line segments generated by noise. An algorithm was written in Python to achieve this (providing similar functionality to the "Extend Line" tool available in ArcMap [33]).
8) All remaining dangles are removed and a polygon topology is built.
9) Finally, polygons outside the agricultural area, or smaller than 1000 m^2, are flagged as invalid. Small areas are likely to represent patches of trees, buildings or other features we do not want to classify. The final vector result is shown in Fig. 4.

When using the field polygons for land-use classification, we use an internally buffered version of the polygons (buffered by 20 m) to extract the per-field spectral statistics. This approach excludes spectrally mixed pixels from around the field edges.

C. Quality Metrics for Segmentation Results

It is not straightforward to judge the "quality" of a segmentation result in a quantitative way. However, there are metrics in the literature for formally assessing a segmentation result (or the line work around the segments) against a set of reference segments, such as a set of hand-delineated fields (e.g., [34]–[36]). Examples include similarity metrics to quantify the average distance between the reference and trial boundaries, or the difference in area between the reference and trial segments.

Locational accuracy is certainly useful for assessing one aspect of correctness, but up to a point it is not the most important aspect for our application of per-field land-use mapping. The more disruptive error type for our purpose is under-segmentation. This is where several "different" fields

are merged into one segment. If our method fails to find the boundary between adjacent fields containing different crops, this can cause an error in the subsequent land-use classification. The spectral and temporal characteristics of the two different land uses will be averaged within the single boundary, so that the result will not be representative of either. The opposite error—over-segmentation—tends to be less serious for our application. If a "real" field is split into several segments, each segment is classified separately, and if all segments fall into the same land-use class then no major error has occurred.

Therefore, we assess the quality of our field boundaries in a two-step process. First, we manually match our field polygons against a set of hand-delineated reference fields in order to assess the cases: under-segmentation, over-segmentation, correct (one-to-one match), and different field layout (many-to-many match). The second step is to assess the locational accuracy of those field boundaries that do have a match with the reference data. The two-step nature of this approach is similar to that by Esch *et al.* [22], where segments are assessed both for the correctness of the segmentation (merging or subdivision of the segments) and then, for the correctly represented image objects, on the basis of deviation in area, perimeter, and shape index.

For our locational accuracy metric, both the field and reference boundaries were converted into raster format, and the locational accuracy of each field boundary pixel was assessed as the distance from it to the nearest reference boundary $d(F_j, R_j)$. The average of these distances is calculated for all N matched boundary pixels, i.e.,

$$\frac{1}{N}\sum_{j=1}^{N} d\left(F_j, R_j\right). \tag{3}$$

This is similar to the metric reviewed in [36], except that we only consider matched boundaries and therefore only need to consider the distance in one direction.

As shown in Fig. 5, we created a set of 273 reference fields (as well as polygons representing other features such as farm yards, buildings, and gardens) by manual delineation in high-resolution QuickBird imagery (KiwImage New Zealand national coverage captured by Digital Globe using the QuickBird satellite, with multispectral imagery at 2.4 m resolution, pan-sharpened to 0.6 m pixels). The imaging date for the northern half of the area was April 7, 2010, and for the southern half it was January 26, 2010 for the panchromatic and February 11, 2010 for the multispectral. These dates fell within the early part of the SPOT satellite time series used in this work.

D. Existing Segmentation Methods for Comparison

For the main comparison of segmentation performance, we used the image segmentation tool in the remote-sensing software package Erdas Imagine 2015. This is an implementation of the multiple-pass region-growing algorithm proposed by Woodcock and Harward [37] but with the addition of an edge detection step, making it an integrated method like [10]. This is an appropriate comparison since 1) it is a standard function available throughout the remote sensing community, and 2) it yields closed segments as ours does, rather than broken raster line work.

Fig. 5. Hand-delineated reference fields, overlaid on the high-resolution satellite imagery from which they were derived (0.6 m pixel size).

We tested Imagine's image segmentation on a subarea from the same dataset used to demonstrate our algorithm. We used the "compute settings" option offered by the tool, which resulted in the following parameters:

1) use Euclidean distance on all 21 layers;
2) apply edge detection with no presmoothing, and a threshold of 6 and a minimal length of 3 pixels;
3) use a minimal value difference of 6 and a variance factor of 3.5 for the segmentation.

We then applied a raster-to-vector conversion (also in Imagine) to the segmentation result.

We also ran the same dataset (seven images with three bands each) through the Spring segmentation algorithm [30]. In common with Hu *et al.* [38], we selected its two parameters by experimenting with a range of values and visually choosing the best result with a similar level of segmentation detail to the other two methods. The parameters used were similarity $= 30$ and area $= 60$. Again the raster result was converted into the vector format.

III. Results and Discussion

A. Study Site and Land-Use Mapping Application

The full area of our main test site, illustrated in Fig. 6, is the Mid-Canterbury plains, South Island, New Zealand, an area of pastoral dairy farming (usually irrigated), sheep and beef farming (usually dryland), and arable cropping. Land use is dynamic, with significant farm conversion over the past 10 to 20 years—mostly to dairy farms. This conversion often involves wholesale removal of existing fences and hedges, installation of large pivot irrigators, and new field layouts. Cropping farms have dynamic crop rotations, with a wide variety of seed, grain, and

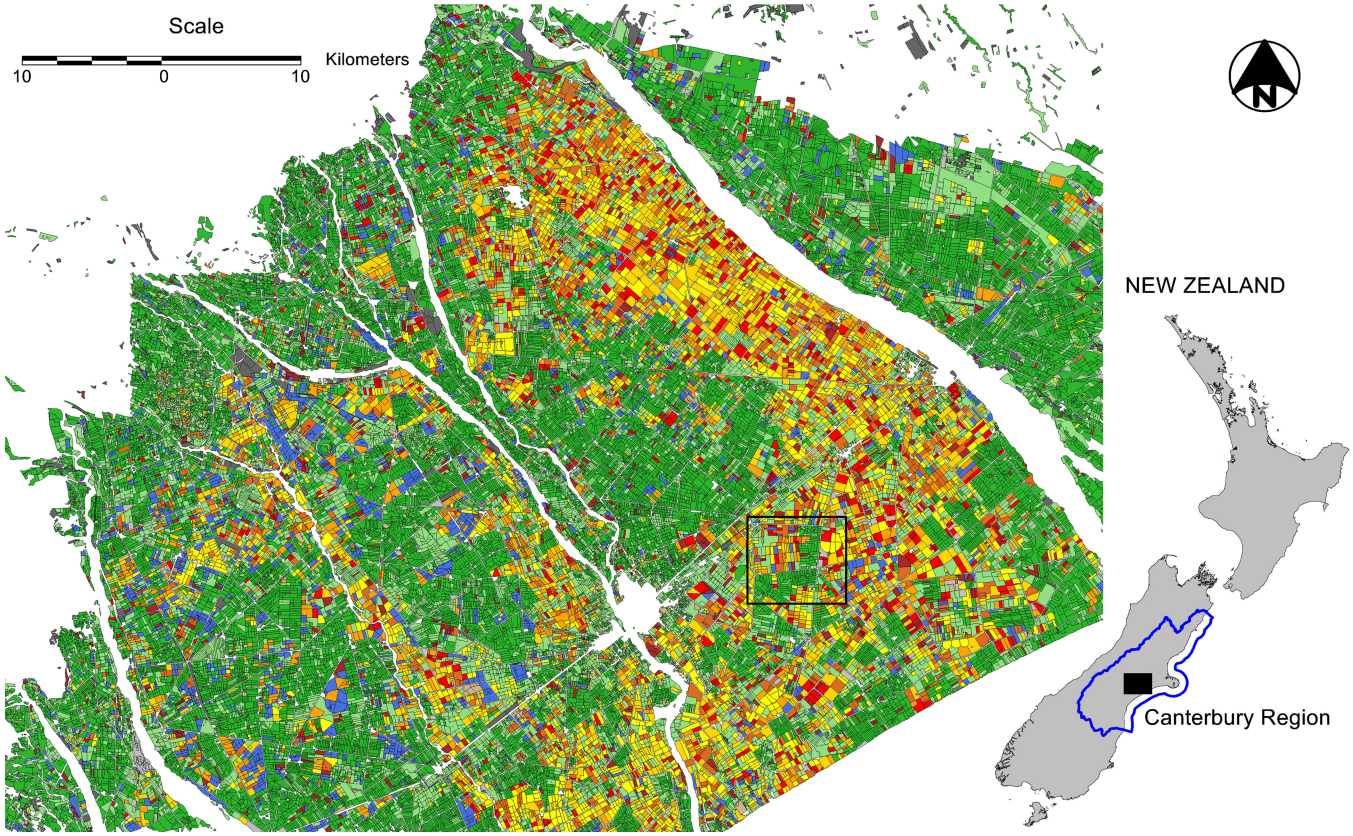


Fig. 6. Study site on Mid-Canterbury plains, South Island, New Zealand, showing one of the land-use classification results that use our field boundary maps as input. Rivers, urban centers, and other nonagricultural areas have been masked out (white). The location of this study site is indicated by a black-filled rectangle on the map of New Zealand. The black-outlined rectangle on the land-use map shows the area of detailed field boundary results presented in Figs. 7 and 11.

vegetable crops grown (some fields will have several different crops per year).

The high degree of land-use and seasonal change has led us to carry out two land-use classifications per year, one centered around a date in summer and a second around a date in winter. Each classification uses a time series of satellite data, typically starting 12 months before and continuing 9 months after the classification date. During this period a time series of some 15 to 19 images may be available. The land-use classifications of the area in Fig. 6 (dated 2010−12) included images from Landsat-5 and -7, and SPOT-4 and -5 [8], and we now also use imagery from Sentinel-2A and -2B for current classifications. Fig. 6 shows one of the summer land-use classification results that uses our field boundary map as input.

To keep up with changes in field layouts, we usually generate a new field boundary map for each classification date, using images from the same period as those used in the classification.

B. Experimental Image Datasets

To demonstrate our method, we operated it on image datasets typical of those we use for land-use mapping. Our field boundary detection method can be applied to any optical imagery. However, of the imagery generally used for the land-use mapping stage, we prefer to select only the higher spatial-resolution SPOT or Sentinel-2 images (typically seven to nine images in the 21-month period) for field boundary mapping.

TABLE I
MAIN MID-CANTERBURY DATASET: IMAGES USED FOR FIELD BOUNDARY DETECTION

Date (UT)	Satellite
18-Feb-10	SPOT-4
7-Mar-10	SPOT-4
5-Sep-10	SPOT-4
2-Oct-10	SPOT-5
18-Jan-11	SPOT-4
15-Feb-11	SPOT-5
21-Aug-11	SPOT-5

The study site in Fig. 6 is just over 4000 km^2 (essentially the area of a full SPOT satellite scene). Each SPOT image was orthorectified to the New Zealand Transverse Mercator map grid at a 10 m pixel size using Imagine Autosync [29]. For the purposes of the land-use classification, all images were radiometrically calibrated for sensor settings and sun angle, though this is not crucial for the field boundary mapping. Cloud masking was carried out as required. The main field boundary results presented in this paper were generated from the set of seven images listed in Table I. For the relatively flat study area in Fig. 6, imagery bands can be used directly. For rolling or hill country we have

TABLE II
DATASETS FOR COMPARISON

	Temuka, Canterbury	Southland	Hawke's Bay
Satellite sensor	Landsat-8	Sentinel-2	SPOT-5
Number of image dates	8	7	4
Pixel size (m)	15	10	10
Image acquisition year	2013	2017	2013–2015
Calibration	Reflectance	Reflectance	Reflectance
Topographic correction	None	Ratios	Ratios

achieved better results by first generating band ratio images as input to the methodology, helping suppress boundaries caused by terrain effects.

To demonstrate the robustness of our method we also include three other datasets as listed in Table II, covering several satellite sensors and varying agricultural landscapes. All these were calibrated to reflectance, compared to our main dataset which used radiance calibration. In all cases, the red, NIR, and SWIR bands were used, sometimes as ratios.

C. Implementation Details—Raster Analysis

Python code was written to perform the steps listed above. This made use of several libraries, including NUMPY, SCIPY, and RIOS [39]. RIOS facilitates automated tiling of an image extent and management of input/output. Although not done for this study, the overlapping tiles can be processed in parallel on different nodes in a cluster computing environment. Processing the seven full SPOT images described in Section III (study site area approximately 4000 km^2) through to raster line work took approximately 25 min on a single computer node. The area was processed in 200 $\times$ 200 pixel tiles. In step 4), tiles have 12 pixels of overlap to accommodate the 13 $\times$ 13 pixel spatial filter. In step 7), tiles have sufficient overlap to accommodate the 300 m minimum fragment size (30 pixels in this case). The per-tile processing means there is no limitation on the spatial extent of the imagery.

The raster analysis steps above are fully automated, and we usually do not vary parameters for a given processing run other than defining the input imagery. If the type of imagery is changed (e.g., sensor, calibration, band ratios), then different band weightings may be required. Other parameters that can be changed are: 1) the threshold on the responses from the directional operators—although we have found our low threshold (10% of the 95th percentile of edge strengths) to be generally appropriate; and 2) the length threshold on connected line fragments to suit local field layout and size—but we find 300 m suitable throughout New Zealand.

D. Implementation Details—GIS Analysis

The GIS steps outlined above are performed on our high-performance computing facility. Processing the approximately 4000 km^2 study area took approximately 2 h to run on a single node with 10 GB of memory.

As with the raster processing, all steps of GIS processing are fully automated, and we usually do not set any parameters other than specifying the input raster line work image and the "agricultural mask" defining the farmland area to be processed. However, parameters that could be varied if required are: 1) allowable extension lengths for GIS dangles; and 2) area of polygons that are considered too small to be a valid field—this will depend on local field sizes and layout, but we find 1000 m^2 suitable for our operational use throughout New Zealand.

E. Generation of Field Boundary Results

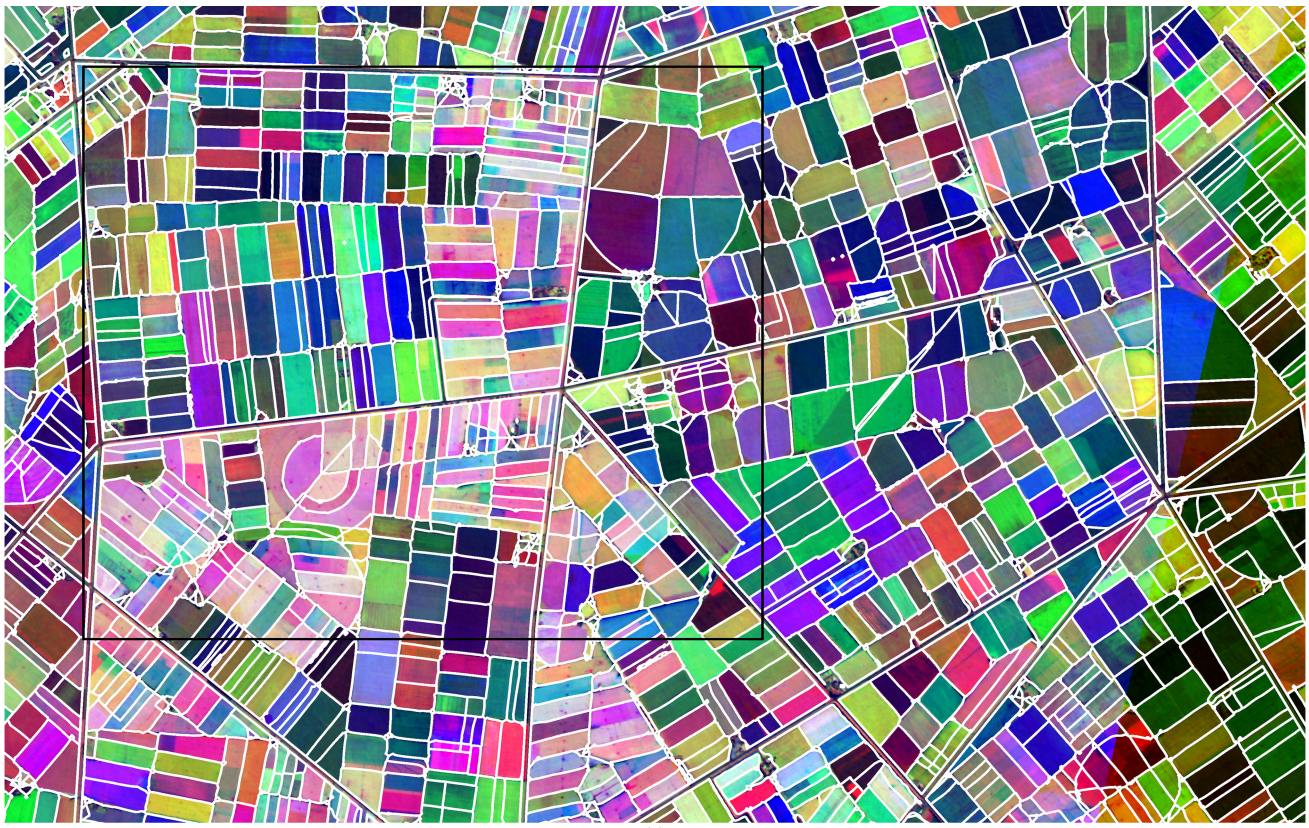
We used our method to automatically segment the full study site shown in Fig. 6 using the seven-date SPOT image dataset listed in Table I. The red, NIR, and SWIR bands of each image were used. The vector result had over 91 000 polygons representing fields. This count excludes a slightly larger number of small polygons that would not be considered as a part of a field, such as isolated buildings, trees, or other small features.

We also applied the Image Segmentation tool in Erdas Imagine to a subarea of the same imagery (1800 $\times$ 1300 pixels) around the reference fields. We noted that the imagery had to be in integer type, not float, for the function to operate successfully. We have successfully trialed the Imagine tool on image sizes up to 8000 $\times$ 8000 pixels (typically yielding 50–60 segments per square kilometer). We presume that the number of segments would finally be limited by the unsigned 32-bit type of the output, though maximum file size or computing resources may impose an earlier limit.

Subareas of both segmentation results are shown in Fig. 7. They are laid over an image composed of three of the image dates, with the NIR band of the three displayed as red–green–blue. The figure gives a good idea of field layout, although further boundaries may be present in dates that are not shown. As a result some field boundaries appear in locations where there is no obvious color difference. Likewise, color differences visible in this three-date combination may not be apparent in some or all of the other dates so may lose significance in the final result, due to the averaging of the STD images. We regard this as a useful feature, as it suppresses boundaries that are transitory (e.g., resulting from a temporary electric fence placed across a field of livestock).

F. Visual Comparison

The immediate visual impression from Fig. 7 is that the field boundaries from our method are much straighter and cleaner than those from the comparison method, though both results appear to successfully segment areas of different color in the composite of imagery. As described in Section II, part of our process is to "burn-in" a buffered version of the road network, so the roads show particularly clearly in our result. This process could, of course, be applied to the result of the Imagine segmentation to achieve a similar effect. Imagine produces a raster result which we converted into vector for display purposes, so it contains pixel staircasing.



(a)

(b)

Fig. 7. Segmentation results for the area around the reference dataset (the area of which is shown by the black rectangle). Line work is shown over an image composed of the NIR band of SPOT images acquired on August 21, 2011, October 2, 2010, and February 18, 2010, with the three dates shown as red–green–blue. (a) Result from our field boundary method. (b) Result from Erdas Imagine's Image Segmentation function.

Fig. 8. Segmentation result from Spring operated on dataset of Table I. Though not visible here, Spring also creates a diagonal line at the edge of one of the images (data gap) as the Imagine algorithm does [see Fig. 7(b)].

The coverage of one of the SPOT images is incomplete along the right-hand side of the subarea. A change in color can be seen in Fig. 7(a) along a diagonal line where the coverage of the February 18, 2010 image finishes, while the other two images have complete coverage. Our method is not disrupted by such data gaps. In comparison, the Imagine function treats these areas as zeros, giving them a different spectral appearance, and therefore putting them into different segments from the areas where there is full coverage on all dates. The diagonal line down the right-hand side of Fig. 7(b) is an artifact of this data gap. This is important, because it is common to have gaps in the data for various reasons, particularly where cloud, haze, and cloud shadow have been masked out.

We also present comparison results as follows.

1) Segmentation from Spring [30] operated on the dataset listed in Table I. The raster result was vectorized for display in Fig. 8. This method produces narrow segments adjacent to boundaries where there is a spectral gradient. Consequently there are many double lines, symptomatic of methods not using edge detection.
2) Application of our method to the other three datasets is listed in Table II. The same parameters were used for all, i.e., $T = 10\%$ of 95th percentile of edge strength, minimum fragment length $= 300$ m, dangle extension up to the lesser of 50% of original length or 200 m. These parameters appear suitable for all three datasets as shown in Fig. 9.
3) Varying our threshold T from 10% to both 5% and 15% of the 95th percentile of edge strength, using the dataset in Table I. These variations are large in comparison to our default threshold, but the effect on the final segmentation is not dramatic (see Fig. 10). More subtle features are delineated with the lower threshold (no spurious noise is introduced) and less with the higher threshold.

Fig. 9. Application of our method to three other datasets listed in Table II. (a) Temuka, Canterbury. (b) Southland. (c) Hawke's Bay. The same parameters were used for all. Line work is laid over three-date NIR images.

Fig. 10. Our method applied with varying threshold T: (a) 5%, (b) 10%, (c) 15% of the 95th percentile of edge strength.

G. Quality Metrics

Simple visual inspection of a segmentation result can provide an indication of the quality of the result. However, the eye can be fooled—the crisp line work looks authoritative, and the human brain finds itself interpreting the image through the frame of the imposed line work rather than judging the line work on its degree of matching to the image.

We compared our main segmentation result and the Imagine result with the reference boundaries to determine matching "units" (minimal matched boundaries). As shown in Fig. 11, cases were:

1) *one-to-one*: where a unit consisted of one segment matching one reference field;
2) *over-segmentation*: where a unit consisted of several segments matching one reference field;
3) *under-segmentation*: where a unit consisted of one segment matching several reference fields; and
4) *many-to-many*: where a unit consisted of several segments matched with several reference fields (different layout).

In each case the unit boundary was considered to be the outer boundary of the group of segments or reference fields of which it was composed.

The unit boundaries were used to assess the locational accuracy, as described in Section II. In both cases, the buffered road network was masked out so that the two results would be comparable. The results are presented in Table III, showing that the average locational accuracy of our field boundaries is slightly better than that of the Imagine result. However, both are very good, with 90% and 88%, respectively, of pixels within 10 m of the true location (10 m is the pixel size of the SPOT satellite imagery). When using the field boundaries for land-use mapping, we internally buffer each polygon by 20 m for calculating the average spectral signature of the polygon. This is to minimize the effect of mixed pixels around the edges. Both segmentation results have 98% of boundary pixels within 20 m of the true location.

As shown in Table III, the mean locational error is around half the 10 m pixel size of the source imagery. When constructing the units for this assessment, we included not only the main, large polygons but also any small polygons that fell with more than half their area within the corresponding reference polygon. This ensured that the locational accuracy assessment was as good as it could be.

However, for the assessment of segmentation shown in Table IV, we counted only the significant polygons, defined to be those larger than 2000 m^2 in size and occupying more than 5% of the unit's area.

Our field boundary matching method has the higher percentage of one-to-one matches (59% as opposed to 39%). The Imagine segmentation has a higher percentage of over-segmented units. This can be seen in Fig. 7, where many fields have line work within them, apparently corresponding to subtle gradients in spectral appearance. These variations could be caused by differences in the health or biomass of crop or pasture, due to, for example, variations in shading, irrigation, soil type, or grazing

Fig. 11. Segmentation results (blue line work) overlaid on reference fields (outlined in red and filled in beige), showing matching between the datasets: (a) our field boundary method and (b) Imagine segmentation. Four units with different matching status are highlighted.

TABLE III
LOCATIONAL ACCURACY METRICS, COMPARED WITH REFERENCE FIELDS

Metric	Field boundary method	Imagine segmentation
Average distance from pixels in segment boundaries to nearest pixel in reference boundaries (m)	5.05 m	5.78 m
Percentage of pixels ≤ 10 m from reference boundary	90%	88%
Percentage of pixels ≤ 20 m from reference boundaries	98%	98%

TABLE IV
SEGMENTATION STATISTICS OF SEGMENTATION RESULTS COMPARED WITH REFERENCE FIELDS

Unit matching between trial and reference boundaries: percentage and number of units	Field boundary method (215 units)	Imagine segmentation (229 units)
One-to-one	59% (127)	39% (89)
Reference field over-segmented	24% (52)	48% (109)
Reference field under-segmented	11% (24)	5% (11)
Different field layout (many-to-many)	6% (12)	9% (20)

intensity. It is not ideal to separate these subtly different parts of fields, as they could then be classified into different land uses. In comparison, our method emphasizes long linear boundaries where STD is relatively high (with lower STD on either side) rather than drawing an arbitrary line between areas of different spectral appearance. The instances of over-segmentation in our field boundary result show larger, straighter partitions, which are more likely to be due to some real difference in treatment of parts of the field—albeit a difference that is not constantly present and was not visible in the single-date reference imagery.

Under-segmented fields are potentially most critical to our use of the field boundary layer for land-use mapping. Under-segmentation could allow statistics from fields with different crops or management to be combined, resulting in a misclassification. In other cases, adjacent fields with identical crops or management (often neighboring pasture fields) may be under-segmented with respect to the reference boundaries, but this would not affect the final land-use classification.

In addition to the results presented in Table IV, we also note that the Imagine result has one unit where a segment is composed of a field inside the reference area and one outside, while our result has two units like this. Also, of the under-segmented and many-to-many units in Table IV, there are two in the Imagine result that include nonfield polygons (e.g., farm yards, gardens), while there are 12 in our result.

Our method is less able to detect boundaries between fields and the typically high-STD areas of farm yards and dwellings amongst them, and this is an area for further work. The Imagine result has many small or narrow polygons that sit across field boundaries—the mixed pixels can form segments in their own right. The preferable result would be for the polygon boundary to lie along the field boundary.

H. Practical Considerations

Our boundary mapping method is viable over large areas: in another project we have processed the whole Canterbury region (outlined in blue in Fig. 6), consisting of 26 000 km^2 of agricultural land within a 360 km × 350 km rectangle. The raster analysis processes the imagery in tiles and could easily be extended to process them in parallel in a high-performance computing cluster environment.

The method is automated, with no complicated parameters to set. We make a careful choice of imagery, preferring a pixel size of around 10 m for suitable representation of field sizes in New Zealand. A good spread of imagery through different seasons is also preferred, to increase the likelihood that boundaries between neighboring fields will be visible on at least some dates. When operating the method in hilly country we use band ratios rather than the original bands, otherwise topographic effects on illumination can create artificial boundaries in the result.

Input imagery often has gaps where cloud or cloud shadow has been masked out, and scenes are variable in extent. Our method ignores no-data areas that are masked or outside image coverage without any disruption to the line work. In comparison, the segmentations of Imagine and Spring treat zeros as data, so no-data areas (in one or more image) will be assigned to a different segment.

We use our method operationally and have run it across several regions of New Zealand, in some cases repeating the analysis using different date ranges and/or satellite sensors.

IV. CONCLUSION

We have developed a segmentation method that is tailored to identify fields in an agricultural landscape. The method has enabled regional-scale analysis of farming patterns by producing field polygons. This allows for the classification of fields as whole objects, which is more accurate than classifying individual pixels. The area that can be segmented is essentially unlimited, because the imagery is tiled for processing.

The locational accuracy of boundary line work is approximately half the pixel size of imagery used to produce it. Comparison with hand-drawn reference boundaries has shown a high degree of segmentation correctness, meaning that the segments seldom merge two different land uses. Both these observations mean that the resulting field boundaries are suitable for input into a land-use classification process.

Comparison against two existing segmentation methods shows straighter, cleaner linework, and a result that is not disrupted by data gaps. We also demonstrated our method on three other New Zealand datasets spanning different regions, topographies, satellite sensors, and radiometric calibrations. In all cases we used the same parameters, showing that our method is robust in the context of the medium resolution satellite imagery we typically use for field scale mapping.

ACKNOWLEDGMENT

The authors would like to thank Z. Hill and J. Cuff (Environment Canterbury) for their support and ongoing

encouragement of work in agricultural land-use mapping, the Manaaki Whenua—Landcare Research Editor, R. Prebble, and our internal reviewer, S. McNeill.

REFERENCES

[1] 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.

[2] R. Lucas, A. Rowlands, A. Brown, S. Keyworth, and P. Bunting, "Rule-based classification of multi-temporal satellite imagery for habitat and agricultural land cover mapping," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, pp. 165–185, 2007.

[3] 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, pp. 290–310, 2007.

[4] T. Esch, A. Metz, M. Marconcini, and M. Keil, "Combined use of multi-seasonal high and medium resolution satellite imagery for parcel-related mapping of cropland and grassland," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 28, pp. 230–237, 2014.

[5] L. R. Lilburne and H. C. North, "Modelling uncertainty of a land management map derived from a time series of satellite images," *Int. J. Remote Sens.*, vol. 31, no. 3, pp. 597–616, Feb. 2010.

[6] H. North, D. Pairman, S. E. Belliss, and J. Cuff, "Classifying agricultural land uses with time series of satellite images," in *Proc. Int. Geosci. Remote Sens. Symp.*, Munich, Germany, 2012, pp. 5693–5696.

[7] H. North, D. Pairman, S. E. Belliss, J. Cuff, and Z. Hill, "Spectral classification of crop groups for land use identification with temporally sparse time-series satellite images," in *Proc. Int. Geosci. Remote Sens. Symp.*, Melbourne, Vic., Australia, 2013, pp. 4237–4240.

[8] H. North, D. Pairman, and S. Belliss, "Full geographic coverage mapping of agricultural land use with time-series satellite images," Landcare Res. Contract Rep. Environ. Canterbury, Christchurch, New Zealand, *Tech. Rep. LC1742*, Jan. 2014.

[9] S. Ullman, *High Level Vision*. Cambridge, MA, USA: MIT Press, 1996.

[10] A. Rydberg and G. Borgefors, "Integrated method for boundary delineation of agricultural fields in multispectral satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 11, pp. 2514–2520, Nov. 2001.

[11] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[12] R. Nevatia and K. Ramesh Babu, "Linear feature extraction and description," *Comput. Graph. Image Process.*, vol. 13, pp. 257–269, 1980.

[13] R. L. Kettig and D. A. Landgrebe, "Classification of multispectral image data by extraction and classification of homogeneous objects," *IEEE Trans. Geosci. Electron.*, vol. GE-14, no. 1, pp. 19–26, Jan. 1976.

[14] S. K. Pal and P. Mitra, "Multispectral image segmentation using the rough-set-initialised EM algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2495–2501, Nov. 2002.

[15] T. V. Robertson, "Extraction and classification of objects in multispectral digital images," in *Proc. Conf. Mach. Process. Remotely Sensed Data*, West Lafayette, IN, USA, Oct. 1973, pp. 27–34.

[16] J. P. Theiler and G. Gisler, "Contiguity-enhanced k-means clustering algorithm for unsupervised multispectral image segmentation," in *Proc. SPIE 3159, Algorithms, Devices Syst. Opt. Inf. Process.*, San Diego, CA, USA, 1997, Paper 3159.

[17] J. Le Moigne and J. C. Tilton, "Refining image segmentation by integration of edge and region data," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 3, pp. 605–615, May 1995.

[18] L. Yan and D. P. Roy, "Automated crop field extraction from multi-temporal web enabled landsat data," *Remote Sens. Environ.*, vol. 144, pp. 42–64, 2014.

[19] J. Graesser and N. Ramankutty, "Detection of cropland field parcels from Landsat imagery," *Remote Sens. Environ.*, vol. 201, pp. 165–180, 2017.

[20] Trimble, *eCognition Developer 8.64: Reference Book*. Munich, Germany: Trimble, 2010. [Online]. Available: http://www.ecognition.com/

[21] U. C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen, "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, pp. 239–258, 2004.

[22] T. Esch, M. Thiel, M. Bock, A. Roth, and S. Dech, "Improvement of image segmentation accuracy based on multiscale optimization procedure," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 463–467, Jul. 2008.

[23] H. C. North, D. Pairman, and S. E. Belliss, "Paddock segmentation using multi-temporal satellite imagery," in *Proc. Int. Geosci. Remote Sens. Symp.*, Québec City, QC, Canada, 2014, pp. 1598–1599.

[24] M. Nagao and T. Matsuyama, "Edge preserving smoothing," *Comput. Graph. Image Process.*, vol. 9, pp. 394–407, 1979.

[25] H. C. North and Q. X. Wu, "An edge-preserving filter for imagery corrupted with multiplicative noise," *Photogramm. Eng. Remote Sens.*, vol. 67, no. 1, pp. 57–64, Jan. 2001.

[26] F. Y. Shih and S. Cheng, "Automatic seeded region growing for color image segmentation," *Image Vision Comput.*, vol. 23, pp. 877–886, 2005.

[27] L. Yan and D. P. Roy, "Conterminous united states crop field size quantification from multi-temporal landsat data," *Remote Sens. Environ.*, vol. 172, pp. 67–86, 2016.

[28] J. A. Long, R. L. Lawrence, M. C. Greenwood, L. Marshall, and P. R. Miller, "Object-orientated crop classification using multitemporal ETM+ SLC-off imagery and random forest," *GISci. Remote Sens.*, vol. 50, no. 4, pp. 418–436, 2013.

[29] Erdas IMAGINE by Hexagon Geospatial. [Online]. Available: http://www.hexagongeospatial.com/products/power-portfolio/erdas-imagine

[30] G. Câmara, R. C. M. Souza, U. M. Freitas, and J. C. P. D. Garrido, "Spring: Integrating remote sensing and GIS by object-oriented data modelling," *Comput. Graph.*, vol. 20, no. 3, pp. 395–403, Jun. 2001.

[31] GRASS GIS. [Online]. Available: https://grass.osgeo.org/. Accessed on: Aug. 2017.

[32] B.-K. Jang and R. T. Chin, "Analysis of thinning algorithms using mathematical morphology," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 6, pp. 541–551, Jun. 1990.

[33] ESRI, ArcMap. [Online]. Available: http://www.esri.com/arcgis/about-arcgis. Accessed on: Aug. 2017.

[34] M. Neubert and H. Herold, "Assessment of remote sensing image segmentation quality," *presented at Geograph. Object-Based Image Anal.*, Calgary, AB, Canada, Aug. 2008. [Online]. Available: http://www.isprs.org/proceedings/XXXVIII/4-C1/

[35] F. Albrecht, S. Lang, and D. Höbling, "Spatial accuracy assessment of object boundaries for object-based image analysis," *presented at Geograph. Object-Based Image Anal.*, Ghent, Belgium, Jul. 2010. [Online]. Available: http://www.isprs.org/proceedings/XXXVIII/4-C7/

[36] R. Q. Feitosa, R. S. Ferreira, C. M Almeida, F. F. Camargo, and G. A. O. P. Costa, "Similarity metrics for genetic adaptation of segmentation parameters," *presented at Geograph. Object-Based Image Anal.*, Ghent, Belgium, Jul. 2010. [Online]. Available: http://www.isprs.org/proceedings/XXXVIII/4-C7/

[37] C. Woodcock and J. Harward, "Nested-hierarchical scene models and image segmentation," *Int. J. Remote Sens.*, vol. 13, no. 16, pp. 3167–3187, 1992.

[38] Y. Hu, J. Chen, D. Pan, and Z. Hao, "Edge-guided image object detection in multiscale segmentation for high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4702–4711, Aug. 2016.

[39] S. Gillingham and N. Flood, RIOS (Raster I/O Simplification). [Online]. Available: http://rioshome.org/en/latest/. Accessed on: Aug. 2017.

Heather C. North (M'03) received the B.Tech. (Hons.) degree in computer technology and the Ph.D. degree in optics/image processing from Massey University, Palmerston North, New Zealand, in 1992 and 1998, respectively.

From 1997 to 2007, she was a Research Scientist with Manaaki Whenua—Landcare Research, Lincoln, New Zealand, specializing in satellite remote sensing. For the past 6 years, she has been a self-employed contractor, with many contracts carried out for Manaaki Whenua—Landcare Research. Her work has focused on agricultural land-use mapping using time-series satellite imagery. Her research interests included speckle suppression and edge detection in synthetic aperture radar imagery, and classification of natural vegetation types in optical imagery.

Dr. North is also a member of the New Zealand Institute of Agricultural and Horticultural Science, and of the Institute of Directors in New Zealand.

David Pairman (M'79) was born in Dunedin, New Zealand, in 1958. He received the B.E. (Hons.) and M.E. degrees in electrical engineering from the University of Canterbury, Christchurch, New Zealand, in 1979 and 1980, respectively, and the D.Phil. degree in image processing in geophysics from Oxford University, Oxford, U.K., in 1986.

From 1980 to 1983, he was a Scientist with the Remote Sensing Section, Physics and Engineering Laboratories, Department of Scientific and Industrial Research (DSIR), New Zealand. After completing the D.Phil. degree, he returned to New Zealand in 1986 to work with the DSIR Division of Information Technology. Since the restructuring of New Zealand science in 1992 he has been working with Manaaki Whenua—Landcare Research, Lincoln, New Zealand, where he leads a group in remote sensing for land-cover and land-use analysis. His research interests include classification techniques, the analysis of synthetic aperture radar imagery, and merging and extraction of information from imagery and GIS data.

Dr. Pairman was the recipient of the National Research Advisory Council (NZ) fellowship in 1983.

Stella E. Belliss was born in Gisborne, New Zealand, in 1952. She received the M.Sc. degree in geology from Canterbury University, Christchurch, New Zealand, in 1978.

From 1977 to 1992, she was with the New Zealand Department of Scientific and Industrial Research in several roles until joining the remote sensing team, where she worked with both optical and SAR imagery. With them, she transitioned to Manaaki Whenua—Landcare Research, Lincoln, New Zealand, a Crown Research Institute, when these were formed in 1992. Her primary role is to engage with client organizations to ensure that Manaaki Whenua—Landcare Research's work is relevant to their operational needs.